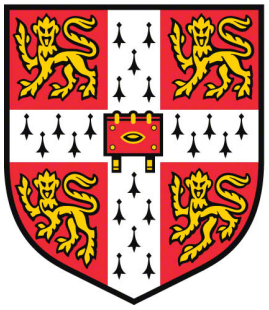
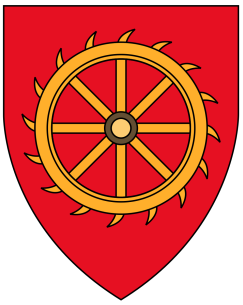
Radiomics and Machine Learning in the Prediction of Cardiovascular Disease





Elizabeth Phuong Vi Le

St. Catharine's College

Division of Cardiovascular Medicine,

Department of Medicine

University of Cambridge

This thesis is submitted for the degree of

Doctor of Philosophy

December 2020

Supervisor: Dr James HF Rudd

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Clinical Medicine and Clinical Veterinary Medicine Degree Committee.

Elizabeth Phuong Vi Le

December 2020

Radiomics and Machine Learning in the Prediction of Cardiovascular Disease

Elizabeth Phuong Vi Le

SUMMARY

Carotid atherosclerosis is a major risk factor for ischaemic stroke which is a leading cause of death worldwide. For stroke survivors, 1 in 4 will have another stroke within five years. Carotid CT angiography (CTA) is commonly performed following an ischaemic stroke or transient ischemic attack to help guide patient management in the secondary prevention of stroke. For example, carotid endarterectomy surgery plus medical therapy or medical therapy alone. The degree of carotid stenosis is the mainstay in making this decision and uses only one aspect of anatomical information that can be obtained from a carotid CTA scan. Radiomics, sometimes called 'texture analysis', is the extraction of quantitative data from medical images that may not be apparent to the naked eye and has already demonstrated clinical utility in oncology for applications ranging from lesion characterisation to tumour grading and prognostication. Machine learning refers to the process of learning from experience (in this case data), rather than following pre-programmed rules.

This thesis presents the findings of a proof-of-principle study to assess the value of radiomics in identifying the 'vulnerable plaque' and the 'vulnerable patient' within the context of cerebrovascular events. To evaluate the potential of radiomic features as imaging biomarkers, their reproducibility and robustness to morphological perturbations were assessed, as well as their biological associations with both PET and immunohistochemistry data. The ability of radiomic features to classify different carotid artery types, namely, culprit, non-culprit and asymptomatic carotid arteries was assessed using several machine learning classifiers. This was subsequently compared with a deep learning approach, which has greater capacity for data mining than feature-based machine learning approaches. Overall, radiomics could extract further useful information from carotid CTA scans. Culprit versus non-culprit carotid arteries in symptomatic patients and asymptomatic carotid arteries from asymptomatic patients had different radiomic profiles that could be leveraged using machine learning for better classification performance than carotid calcification or carotid PET imaging alone. Reliable and robust CT-based carotid radiomic features were identified that were associated with the degree of inflammation underlying the carotid artery. If validated with future prospective studies, this has the potential to improve personalised patient care in stroke management and advance clinical decision-making.

Acknowledgements

I am grateful to the MB PhD programme and my funders: the Cambridge School of Clinical Medicine, the Medical Research Council's Doctoral Training Partnership and the Frank Edward Elmore Fund for providing me with the opportunity to explore the world of academic research and to develop new skills over the last three years. This would not have been possible without my supervisor, Dr James Rudd, who gave me the opportunity, even when I was a third-year medical student, to get involved with summer research projects in the Department of Medicine and who believed in me when I applied to the programme. I am extremely grateful for his continuous support, mentorship and guidance. I would like to extend my sincere thanks to my second supervisor, Professor Fiona Gilbert, for her support and for allowing me to learn more about the application of machine learning to breast imaging.

My understanding of machine learning and biomedical image analysis greatly benefitted from my time at Imperial College London and the BioMedIA team. I am deeply grateful to Professor Carola-Bibiane Schönlieb and the Centre for Mathematical Imaging in Healthcare Partnership Fund for providing a travel grant to facilitate this experience and to Dr Ben Glocker and his colleagues and students for giving up their time to mentor me, answer my questions and include me in their weekly journal clubs.

I would like to thank Dr Nicholas Evans, Dr Jason Tarkin, Dr Francis Joshi and Dr Mohammed Chowdhury for their permission to use their vascular imaging datasets for my research, as well as for their kind welcome and support when I joined the team. Without the guidance of Dr Nicholas Evans and Dr Jason Tarkin from the very beginning and throughout my PhD, this thesis would not have been possible. I am also grateful to Dr Fulvio Zaccagna for teaching me about interpreting carotid scans and training me to use TexRAD, along with Dr Balaji Ganeshan. I wish to thank Dr Jonathan Weir-McCall for teaching me about cardiac CT imaging, carotid luminal stenosis measurements and for always encouraging me following our discussions. I would also like to extend my thanks to Dr Deepa Gopalan for her patience and enthusiasm in teaching me about cardiac imaging.

I am deeply indebted to Holly Pavey and Dr Yuan Huang for their invaluable advice about statistical methodology and to Dr Leonardo Rundo, whose unrelenting optimism helped me greatly in writing this thesis throughout the pandemic and whose insightful suggestions taught

me more about radiomics and robustness analyses. I am also grateful to Dr Michael Roberts for his feedback, which has taught me more about radiomics and machine learning. I would like to thank Dr Chris Wall for assisting me by being the second rater in the reproducibility studies of this thesis and to thank Dr Rouchelle Sriranjan for helping me with determining the clinical characteristics of the carotid arteries i.e. carotid plaque type. I would like to acknowledge the assistance of Parvesh Konda who quantified the %CD68+ from immunohistochemistry slides as part of his SSC with the Department of Cardiovascular Medicine.

In addition, my special thanks go to my friends, in particular Matt Judge for introducing me to the basics of programming and Mica Ménard for his continuous encouragement and support. My MB PhD colleagues, Joachim Hanna and Maria Fala, have become my Cambridge family and I am truly blessed to have them in my life. I am so thankful that we are able to support each other through all the trials and tribulations of clinical medicine, PhD life and beyond. Finally, I owe my deepest gratitude to my parents and my brother for their unwavering love, support and for being there for me every step of the way.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ACAS	Asymptomatic Carotid Atherosclerosis Study
ACC/AHA	American Heart Association/American College of Cardiology
ACS	Acute coronary syndrome
ACST	Asymptomatic Carotid Artery Surgery Trial
AI	Artificial intelligence
ANN	Artificial neural network
Asx	Asymptomatic carotid artery
AU	Agatston units
AUC	Area under the receiver operating characteristic curve
b	Bias
BH	Benjamini-Hochberg correction
BioMedIA	Biomedical Image Analysis Group at Imperial College London
BN	Bin number
BW	Bin width
CAC	Coronary artery calcium score/ Carotid calcium score
CAD	Coronary artery disease
CAS	Carotid arterial stenting
CC	Culprit carotid artery
CEA	Carotid endarterectomy
CHAI	The Carotid Hypoxia and Inflammation study
CHD	Coronary heart disease
CI	Confidence interval
CLI	Command line interface
CNN	Convolutional Neural Network
CNR	Contrast-to-noise ratio

COVID-19	Coronavirus Disease 2019
CPU	Central processing unit
CREST	The Carotid Revascularization Endarterectomy versus Stenting Trial
CT	Computed tomography
CTA	Computed tomography angiography
CVD	Cardiovascular disease
DAB	3,3'-diaminobenzidine
DC	Dice coefficient
DICOM	Digital Imaging and Communications in Medicine
DL	Deep learning
DOTATATE	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid –[Tyr3]octreotate
DPD	Dry pleural dissemination
DT	Decision Tree
ECA	External carotid artery
ECG	Electrocardiography
ECST	European Carotid Surgery Trial
FDG	Fluorodeoxyglucose
FMISO	Fluoromisonidazole
FN	False negatives
FP	False positives
FRP	Fat radiomic profile
GLCM	Grey level co-occurrence matrix
GLDM	Grey level dependence matrix
GLRLM	Grey level run length matrix
GLSZM	Grey level size zone matrix
Grad-CAM	Gradient-weighted Class Activation Mapping
GUI	Graphical user interface
HCM	Hypertrophic cardiomyopathy
HHD	Hypertensive heart disease

HU	Hounsfield units
IBSI	Imaging Biomarker Standardisation Initiative
ICA	Internal carotid artery
ICARUSS	The Imaging Carotid Atherosclerosis in the Recovery and Understanding of Stroke Severity Study
ICC	Intraclass correlation coefficient
IH	Intensity histogram
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IQ	Image quality score
IQR	Interquartile range
IVUS	Intravascular ultrasound
KNN	K-nearest neighbours
LASSO	Least absolute shrinkage and selection operator
LDL	Low-density lipoprotein
LoG	Laplacian of Gaussian
LR	Logistic regression
MACE	Major adverse cardiovascular events
MESA	Multi-Ethnic Study of Atherosclerosis
MI	Myocardial infarction
ML	Machine Learning
MPP	Mean of positive pixels
MRI	Magnetic resonance imaging
MS	Multi-slice
NaF	Sodium fluoride
NASCET	North American Symptomatic Carotid Endarterectomy Trial
NC	Non-culprit carotid artery
NGTDM	Neighbouring grey tone difference matrix
NICE	National Institute for Health and Care Excellence
NifTI	Neuroimaging Informatics Technology Initiative

NN	Neural network
NRS	Napkin-ring sign
NSCLC	Non-small cell lung cancer
PET	Positron emission tomography
PVAT	Perivascular adipose tissue
PVO	Prosthetic valve obstruction
RAM	Random access memory
ReLU	Rectified linear activation function
RF	Random forest
RGB	Red-green-blue
RMSProp	Root Mean Square Propagation
ROI	Region-of-interest
r_s	Spearman's rho
SCOT-HEART	Scottish Computed Tomography of the Heart trial
SD	Standard deviation
SHAP	Shapley Additive Explanations
SNR	Signal-to-noise ratio
SSF	Spatial scale filter
SUVmax	Maximum standardized uptake values
SVM	Support vector machine
TBRmax	Maximum tissue-to-background ratio
TIA	Transient ischaemic attack
TL	Transfer learning
TN	True negatives
TNF	Tumour necrosis factor
TP	True positives
US	Ultrasound
VEGF	Vascular endothelial growth factor
VGG	Visual Geometry Group

VISION	The Vascular Imaging Inflammation imaging using Somatostatin receptor positron emission tomography study
VOI	Volume-of-interest
w	Weight
XML	Extensible Markup Language
YAML	YAML Ain’t Markup Language

Chapter 1: Introduction and Background

Chapter Summary:

The aim of this chapter is to provide the relevant background information regarding cardiovascular disease, cardiovascular imaging, radiomics and machine learning for this research project. The current limitations and challenges of radiomics and machine learning are discussed with respect to their clinical applications and a review of the literature as well as the common terminology relevant to these fields are presented.

1.1 Cardiovascular Disease

1.1.1 Epidemiology of Cardiovascular Disease

Cardiovascular disease (CVD) is an umbrella term that refers to diseases of the heart and circulatory system, including acute events such as heart attacks and strokes. Although CVD mortality in the UK has been declining over the last few years[1], CVD remains the leading cause of death both nationally and internationally[2,3], accounting for 31% of all deaths in the world[4]. 1 in 4 deaths are attributable to CVD in the UK alone, equating to 1 death occurring every 4 minutes[4] with an estimated societal cost of £26 billion per year[5]. In addition, patients with CVD have an increased risk of other chronic diseases and comorbidities[6] such as rheumatoid arthritis, diabetes mellitus and gout. More recently, amidst the COVID-19 pandemic, CVD was found to be a risk factor for a poorer prognosis of COVID-19[7,8]. Taken together, CVD represents a significant health and economic burden to society in terms of mortality and morbidity.

1.1.2 Pathophysiology of Atherosclerosis

Atherosclerosis is the common pathophysiological process underlying coronary artery disease (CAD) and carotid artery disease. Characterised by lipid accumulation in the arterial wall leading to the formation of a plaque (atheroma), this chronic inflammatory process has varying stages of severity and can affect different vascular beds from large arteries such as the aorta to smaller peripheral arteries such as the femoral arteries. Acute cardiovascular events such as myocardial infarction (MI) and ischaemic stroke arise when these atherosclerotic plaques are

disrupted (typically through plaque rupture rather than plaque erosion)[9] and thrombosis or vessel occlusion occurs[10].

Plaque rupture, as opposed to plaque erosion, refers to a defect in the fibrous cap of the plaque which leads to exposure of its thrombogenic core. In plaque erosion, the fibrous cap remains intact but disruption of the endothelium results in thrombosis. Although the prevalence of plaque erosion as the culprit mechanism for acute coronary syndromes (ACS) is gaining more attention[11], plaque rupture is still the predominant causal mechanism of coronary thrombosis[9].

Traditional cardiovascular risk factors such as hypertension, hyperlipidaemia, smoking and diabetes mellitus are thought to drive the progression of atherosclerosis by inducing endothelial damage. As a consequence of endothelial damage, low-density lipoproteins (LDLs) are oxidized and deposited within the vessel wall which triggers a cascade of inflammatory processes leading to adaptive changes in the vessel wall and atherosclerotic plaque progression.

Most plaques cause subclinical disease and the patient remains asymptomatic. They tend to have a thick, highly calcified, fibrous cap that can better shield the thrombogenic lipid core from the lumen of the vessel. These rarely cause thrombosis but can sometimes produce ischaemic symptoms such as angina due to luminal narrowing and a reduced ability to match oxygen supply to oxygen demand for the heart muscle cells[12].

A ‘vulnerable’ or unstable high-risk plaque may trigger a cardiovascular event through cap rupture or erosion. Histologically, high-risk plaques have the following features[13], illustrated in Figure 1.1.

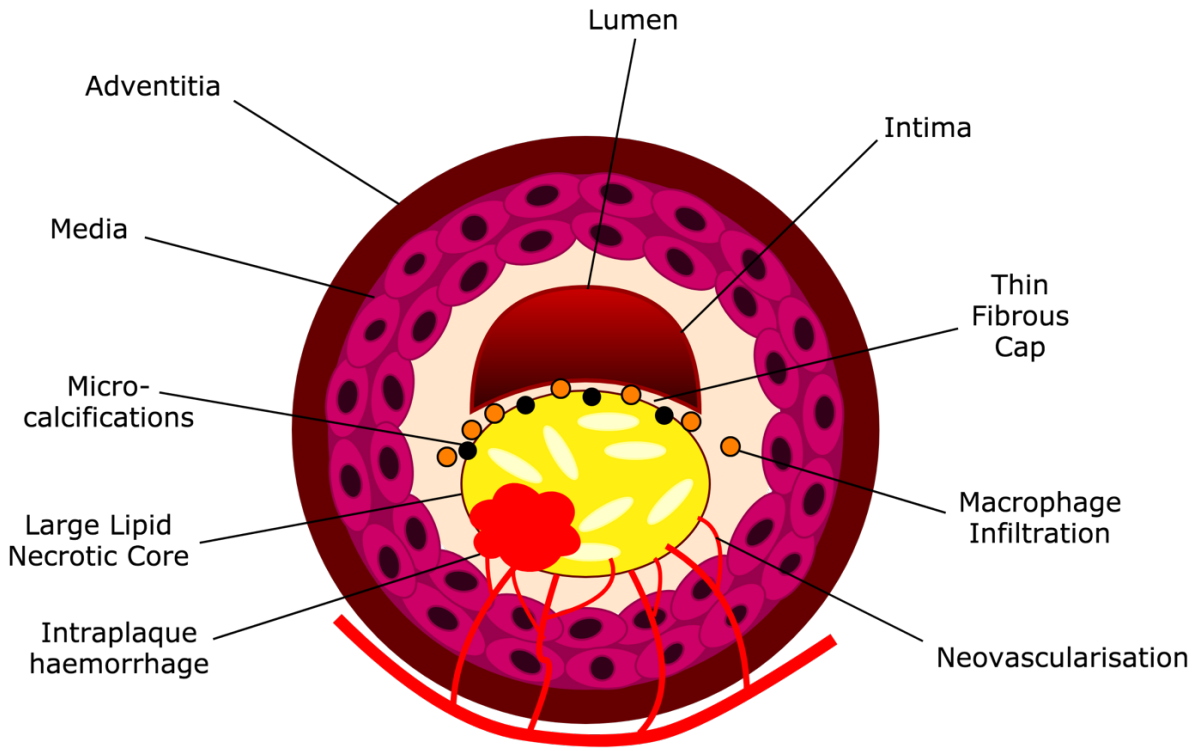


Figure 1.1 The 'vulnerable' plaque

The presence of a thin fibrous cap, a lipid-rich necrotic core, macrophage infiltration, intraplaque haemorrhage and neovascularisation and microcalcification processes are histological characteristics of a vulnerable plaque.

Although plaque expansion may lead to luminal encroachment and stenosis, high-risk plaques are associated with positive (outward) remodelling of the arterial segment which may spare luminal patency and prevent obstruction to blood flow[13]. This explains why the severity of pre-existent stenosis is not a reliable predictor of future acute plaque events. Furthermore, a high burden of disease may be present in the absence of significant stenosis[14].

The ultimate fate of different atherosclerotic lesions and the speed of their progression is difficult to predict. Local factors relating to the atherosclerotic plaque such as low endothelial shear stress and structural features such as vessel anatomy may contribute to plaque instability, but systemic factors may also play a role including infection and autoimmunity[15]. Genetic factors and lifestyle risk factors have been identified as independent contributors to the risk of developing CAD, but the extent of their contribution amongst different individuals is difficult to determine[14].

1.2 Cerebrovascular disease

1.2.1 Transient Ischaemic Attacks and Stroke Definitions

Every two seconds, someone in the world will have a stroke[16]. In the UK, there are more than 100,000 strokes every year with around 1.2 million people in the UK living with the consequences[17]. In contrast to ischaemic heart disease, stroke is more heterogeneous in nature with several aetiological subtypes. An updated definition of stroke was provided by the American Heart Association/American Stroke Association in 2013 with ischaemic stroke referring to "an episode of neurological dysfunction caused by focal cerebral, spinal or retinal infarction on imaging studies"[18]. On a spectrum of severity, transient ischaemic attacks (TIA) are defined as brief episodes of neurological dysfunction following focal cerebral ischaemia without permanent cerebral infarction[19].

Extracranial carotid artery atherosclerosis, also known as carotid artery disease, accounts for approximately 20% of all ischaemic stroke[20] and TIA episodes. Providing supporting evidence that atherosclerosis is a systemic disease, carotid artery disease often occurs concomitantly with coronary heart disease and the former is independently related to an increased risk of cardiovascular events[21].

Carotid plaque rupture may lead to distal embolization of thrombus or complete occlusion of the artery. Either outcome results in compromised cerebral perfusion and subsequent TIA or stroke depending on factors such as collateral compensation.

1.2.2 Carotid Anatomy

In terms of carotid anatomy, the right and left common carotid arteries have different origins. Whereas the left common carotid artery originates directly from the arch of the aorta, the right common carotid artery arises from the bifurcation of the brachiocephalic trunk[22]. Nevertheless, both common carotid arteries extend into the neck and divide into the external and internal carotid arteries typically at the level of the superior margin of the thyroid cartilage, see Figure 1.2. This area is referred to as the carotid bifurcation. Just superior to the carotid bifurcation is a dilated area known as the carotid bulb which is the location of the carotid sinus with baroreceptors important for blood pressure regulation[23].

Carotid artery plaque is most frequently found at the carotid bifurcation, often with carotid atherosclerosis extending into the area of the carotid bulb to involve the proximal internal carotid artery (ICA) rather than the common carotid artery[24]. Approximately 10 to 12 percent of all ischaemic strokes are attributed to atherosclerosis of the ICA at the level of the carotid bifurcation[25,26]. This predisposition of carotid atherosclerosis occurring at the carotid bifurcation may relate to the geometry of the carotid artery and its associated fluid haemodynamics, since the carotid bifurcation is associated with low wall shear stress (where shear stress refers to the lateral biomechanical force of blood flow experienced by the vessel endothelium[27]), resulting in flow stagnation[28].

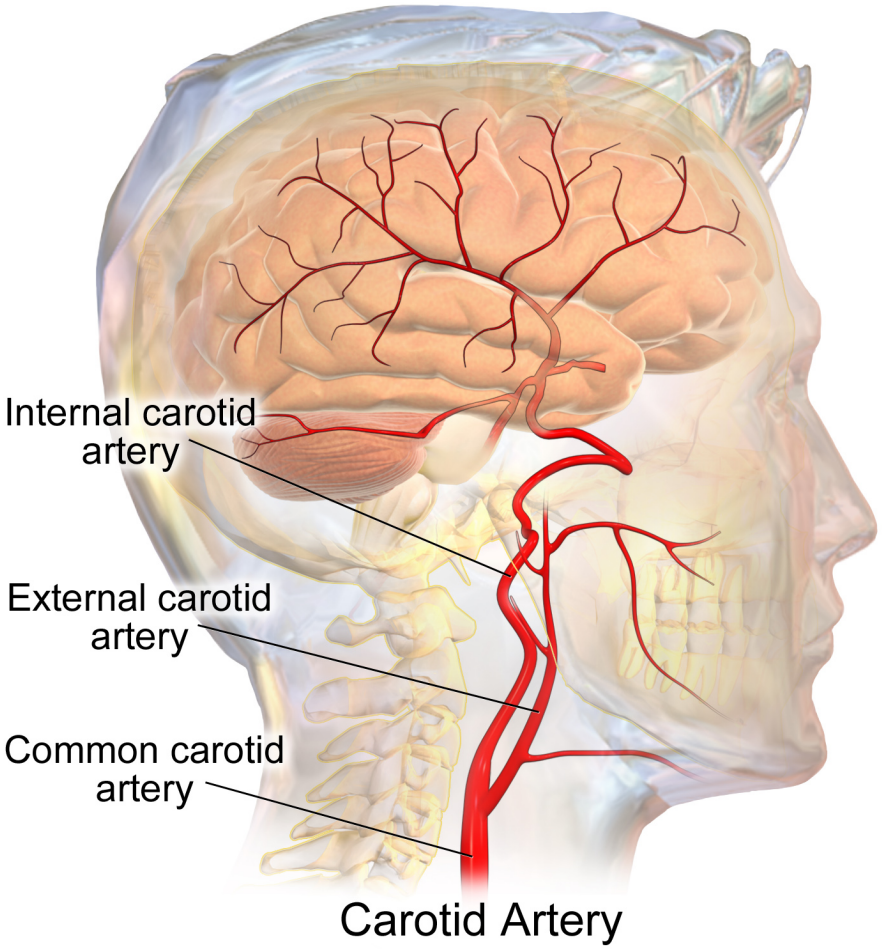


Figure 1.2 Carotid artery anatomy

Reproduced from Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436[29].

1.2.3 Risk of Recurrent Cerebrovascular Events

Following a TIA, there is substantial risk of a subsequent stroke. The incidence of prior TIA in patients with stroke is approximately 15-30%[30]. Data analysis of the Oxford Community Stroke Project suggests that the short-term stroke risk from onset of first TIA is 8.6% (95% CI, 4.8 to 12.4) within 7 days and 12.0% (95% CI, 7.6 to 16.4) within 30 days[31]. For stroke survivors, around 1 in 4 will have another stroke within five years, with the greatest risk being within the initial thirty days following the first event[32]. In either case, there is a critical window of opportunity for diagnostic imaging of the unstable culprit plaque (i.e. the plaque that had precipitated the stroke/TIA and which may lead to a subsequent recurrent event without intervention) following the cerebrovascular event, before appropriate therapeutic interventions have been delivered to enable a closer study of the 'vulnerable plaque'.

1.2.4 Carotid Intervention for Stroke Prevention

In the context of carotid atherosclerosis, symptomatic patients refer to individuals who have previously experienced an ischaemic stroke or TIA, whereas asymptomatic patients have never previously experienced a cerebrovascular event. Secondary prevention refers to therapies or interventions administered to symptomatic patients in order to reduce the risk of a recurrent event. In the 1990s, large randomised trials (NASCET[33] and ECST[34]) demonstrated the effectiveness of carotid endarterectomy (CEA) in comparison with conventional medical treatment for reducing recurrent stroke risk. A pooled analysis of the data found that CEA was particularly beneficial for symptomatic patients with 70% to 99% carotid stenosis[35], but this also depended upon the patient being surgically-fit and the time interval between the cerebrovascular event and the surgery. More recently, evidence suggests that carotid artery angioplasty and stenting (CAS) could be non-inferior alternatives to CEA[36,37]. In either case, the degree of carotid artery stenosis, which is typically expressed as the percentage reduction in vessel diameter as measured from non-invasive imaging (such as carotid ultrasound or CT angiography), forms an important part of the patient selection criteria.

In light of more effective medical therapy, the effectiveness and appropriateness of carotid revascularisation for asymptomatic patients with carotid stenosis is less clear than for those with symptoms[38]. Whilst the ACST[39] and ACAS trials[40] for asymptomatic carotid stenosis (70%-99% in ACST and 60%-99% carotid stenosis patients in ACAS) demonstrated a small

benefit in reducing the risk of ipsilateral stroke with surgery, these trials were conducted in the 1980/90s prior to current medical therapies. In addition, carotid revascularisation is associated with its own risks. The CREST trial found that there was a 2.5% risk of stroke or death following carotid stenting and a 1.4% risk following endarterectomy within 30-days following the procedure[36,38]. Therefore, in the UK, the National Institute for Health and Care Excellence (NICE) guidelines for the management of asymptomatic extracranial carotid stenosis do not recommend routine carotid revascularisation for primary stroke prevention, but the guidelines encourage clinicians to consider recruiting patients to the Asymptomatic Carotid Artery Surgery Trial 2[41] to address this uncertainty[42].

Even though the degree of carotid artery stenosis is related to the risk of stroke[43], the majority of asymptomatic patients with carotid stenosis will remain free from cerebrovascular events, even without surgical intervention[44]. Furthermore, whilst the criteria for carotid revascularisation focuses mainly on 70% carotid stenosis (classified as high-grade carotid stenosis), the prevalence of these lesions in the asymptomatic population ranges from zero to 1.6%[45] and the prevalence in the symptomatic population is less than 5%[46]. As such, the degree of carotid artery stenosis alone fails to adequately identify 'vulnerable plaques' and 'vulnerable patients'. We need better tools to characterise stroke risk.

1.2.5 Carotid Artery Imaging and Evaluation

With the development and refinement of different imaging modalities, it is possible for researchers to move beyond luminal stenosis and incorporate non-invasive plaque imaging information to improve patient risk stratification[47]. Histopathological studies have shown that particular characteristics of carotid plaque composition and morphology are associated with increased susceptibility to rupture, independent of arterial stenosis. These so-called 'hallmarks' of instability include a thin fibrous cap, large lipid core, presence of intraplaque haemorrhage and plaque ulceration[48] and are similar to coronary plaque instability characteristics shown in Figure 1.1.

A systematic review found that MRI was effective at identifying and evaluating carotid plaque characteristics such as a lipid-rich necrotic core and that these features were associated with an increased risk of stroke or TIA[49]. However, carotid CT is the gold standard for detecting and quantifying carotid calcification. CT angiography (CTA), whereby an iodine-based contrast agent is injected to visualise the carotid artery lumen, is a robust method for assessing the degree of carotid stenosis[50] and can be used to characterise plaque type such as soft plaques, fatty plaques and calcified plaques according to attenuation level[51,52].

In the management of stroke or TIA affecting the anterior cerebral circulation, a carotid Doppler ultrasound is typically the first-line test for diagnosis of carotid artery disease, with CTA follow-up for further evaluation of carotid plaque if necessary. Sometimes MR angiography may be used instead according to local trust guidelines, however since CTA is more widely available, this is the more commonly used modality in the UK[53].

1.2.6 Coronary and Carotid Calcium scoring

Coronary artery calcification is typically quantified using the Agatston score[54], a method that uses a multiplication factor for ranges of calcium attenuation values and factors in the area of the calcified plaque. The coronary artery calcium score (CAC) provides a standardised method for quantifying arterial calcification as a surrogate marker of atherosclerosis[55]. Studies have shown that the CAC score provides important prognostic information beyond that derived from traditional risk factor scoring alone[56]. However, the US Preventive Services Task Force currently does not recommend CAC (or any other test) for CVD screening in asymptomatic adults due to a lack of evidence of its efficacy[57], but it can be used for risk refinement.

A high CAC (CAC>100 or CAC>75th percentile for age and gender) shifts conventional risk estimations upwards, whilst a zero CAC score shifts it downwards. This rationale relates to the increased cardiovascular event rates associated with an increased absolute CAC score, as demonstrated in Figure 1.3. A zero CAC is associated with a good prognosis and has a high negative predictive value for excluding obstructive coronary artery disease, however, there remains a small risk of non-calcified plaques[58].

Because calcification is a widely used indicator of atherosclerosis, there has been increased interest in CT quantification of carotid calcification. However, the association of carotid calcification and stroke risk is currently unclear and the methods for carotid calcification quantification are non-standardized[59]. Methods vary from semi-quantitative measures of calcification such as *moderate/heavy vs none/minor calcification* to quantitative measures such as *calcification volume* or carotid calcium scoring adapted from the Agatston-based method used for coronary artery calcium scoring[54,59]. A meta-analysis in 2010 suggested that culprit plaques have a lower degree of calcification than non-culprit plaques, with the caveat that there were great heterogeneity in the methodology to quantify carotid calcification between the 24 studies included in that review[59].

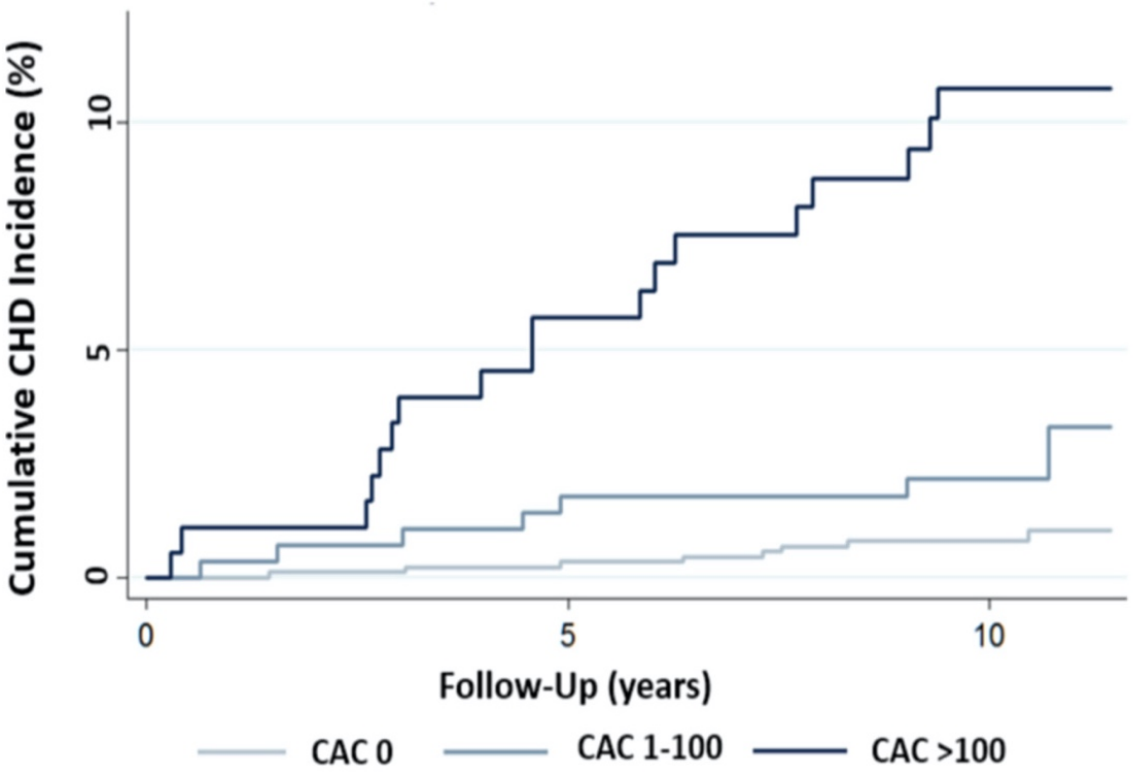


Figure 1.3 Events rates and CAC category.

The cumulative incidence of coronary heart disease (CHD) in the MESA low lifetime risk sample increases with increasing burden of subclinical atherosclerosis as determined by CAC category (CAC 0, >0, >100). Reproduced from Joshi et al. Atherosclerosis 246, 367-373 (2016)[60].

There remains a need to clarify the role of carotid calcification and risk of stroke in asymptomatic and symptomatic patients, using a quantitative and reproducible method. In addition, the information within CT and CTA scans is not limited to carotid luminal stenosis or carotid calcification. With radiomics, we can quantitatively exploit the information embedded within medical images in a more sophisticated way than simply measuring the degree of luminal stenosis, and this may aid in the discovery of imaging biomarkers for better risk prediction in stroke and TIA.

1.3 Radiomics

Radiomics, sometimes called 'texture analysis', refers to the extraction and analysis of quantitative data[61] from medical images that may not be apparent to the naked eye[62]. Already an established field of research in oncology for tumour characterisation, radiomics derives quantitative features from non-invasive imaging modalities such as CT, MRI and PET for developing diagnostic and prognostic prediction models[63,64]. A potential benefit of quantitative medical image analysis is the ability to bypass visual interpretation subjectivity as well as the ability to combine these radiomic features with non-imaging data to improve clinical decision making.

Radiomics can extract several thousands of features which cover a wide range of quantitative attributes that are mathematical descriptions of the visual properties of an image. They are human 'engineered' and can be grouped into different categories[65] as shown in Table 1.1. Because radiomic features are derived from existing patient imaging datasets, these features can provide valuable information for personalised medicine in a fast, cost-effective and non-invasive manner, without the need for extra testing.

Prescott et al. provide a definition for a quantitative imaging biomarker as 'an imaged characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or a response to a therapeutic intervention."[66] In oncology, several radiomic features have been correlated with tumour pathology such as hypoxia[67] and angiogenesis[68]. Whilst little is known about vascular tissue radiomics, we know that inflammation is a common pathway in tumour development and atherosclerosis progression[69]. Atherosclerotic plaque hypoxia and angiogenesis relates to increased risk of rupture[70] and this provides the rationale for investigating the radiomics of carotid atherosclerosis. Through similarities between tumour pathophysiology and atherosclerosis, radiomic features derived from carotid imaging may provide relevant quantitative imaging biomarkers for cardiology and stroke medicine.

Table 1.1 Different Radiomic Feature Categories

Features	**Definition**
First Order statistics	Histogram-based methods e.g. extraction of features from the histogram of voxel intensities i.e. Hounsfield units for CT images. These include the mean value, maximum and minimum, standard deviation (dispersion), skewness (asymmetry) and kurtosis (sharpness/flatness). These features do not consider the spatial relationships between pixels, unlike higher-order statistics.
Higher-order statistics	Greyscale variation features summarize the local spatial inter-relationships within a region. The grey level co-occurrence matrix (GLCM) consists of counts for voxel pairs with certain grey values at a predefined direction and distance from each other. Features generated from this matrix include similar/dissimilar contrast values.
Wavelet decomposition	Wavelet decomposition of the image extracts intensity and textural features from different frequency bands. This method can be used to obtain fused texture characteristics from two imaging modalities.
Structural/shape-based features	Describes the 3D geometrical composition of the segmented structure and includes sphericity, compactness and surface-to-volume ratio.

1.3.1 Radiomics Workflow

There are several stages to the radiomics workflow, illustrated in Figure 1.4. The initial stages involve image acquisition and image reconstruction to obtain the medical images to be analysed[71]. Subsequently, the structures of interest within the image such as a tumour or the carotid artery need to be delineated to identify the regions-of-interest within the image (ROI) – the area(s) of the image from which radiomic features will be extracted. This is known as segmentation of the image. This may be done manually, a time-consuming process subject to intra- and inter-observer variability, via a semi-automatic approach or using a fully automated segmentation method[72].

Delineating a structure on a single slice of the imaging dataset results in an ROI. It is possible to draw ROIs on several consecutive image slices and concatenate these ROIs together to form a volume-of-interest (VOI) in order to extract 3D radiomic features, i.e. radiomic features that consider the grey level values within the x, y, and z dimensions of the dataset[73]. Prior to radiomic feature extraction, image pre-processing may be applied to the images such as normalisation or resampling the image to isotropic values[71]. This is followed by image quantisation which converts the image grey level values into a discrete set of grey level counts via fixed bin number or fixed bin size discretisation methods (discussed further in Chapter 3). This step is necessary to enable radiomic feature calculations[73] from the segmented regions. Features are extracted from the ROI/VOIs across several radiomic feature classes (also referred to as categories, see Table 1.1), that are available in a radiomics software package or coded in-house. These radiomic features can then be used as the input for a machine learning model or analysed with traditional statistical methods.

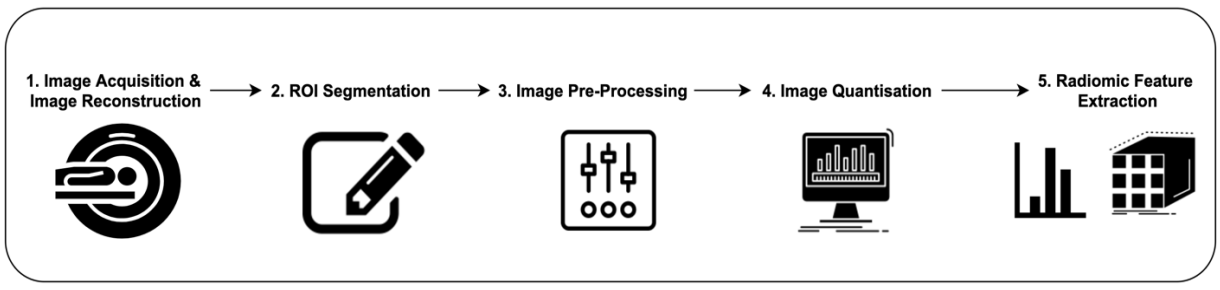


Figure 1.4 Radiomics Workflow

1.3.2 Radiomics Software Packages and Applications

There are several open-source and commercial software packages available for radiomics analysis, including the option to program in-house radiomics tools using programming languages such as MATLAB, R and Python. In this thesis, the commercially available research texture analysis software TexRAD, and the open-source Python package PyRadiomics were used and will be further described below.

1.3.2.1 TexRAD: a commercial research software for texture analysis

TexRAD (Feedback Medical Ltd, Cambridge, UK) is a commercial research texture analysis software. Unlike other radiomics packages, TexRAD extracts a relatively limited number of textural features (approximately 30). This set of pre-determined first order statistics is extracted from the region-of-interest by using a filtration-histogram approach. This involves image filtration with a Laplacian of Gaussian (LoG) band-pass filter to reduce image noise and to highlight image features of a specified size at various scales (fine, medium and coarse features in tumour radiomics) by altering the width of the LoG filter, followed by quantification of these features derived from the image histogram[74].

Texture features can also be extracted from the original unfiltered image. Derived features include properties of the grey level histogram of the ROI, such as the first order histogram standard deviation, kurtosis, entropy, mean and skewness, see Figure 1.5 and Figure 1.6. First Order statistics discard spatial information and refer to parameters which describe different aspects of the distribution of grey level values in the area of interest[68]. In the case of CT images, this would be different properties of the distribution of Hounsfield units in the ROI, see Figure 1.5.

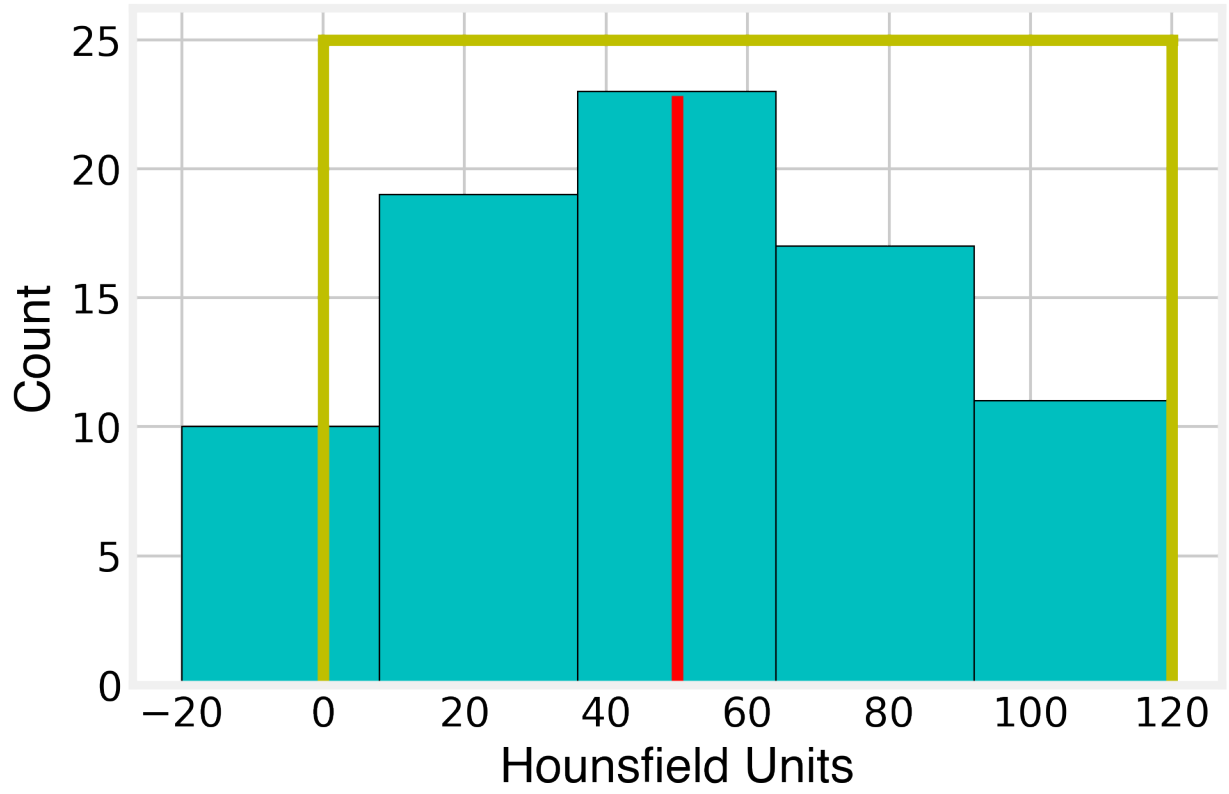


Figure 1.5 First Order statistical features derived from CT

The x-axis represents grey level values or attenuation, and the y-axis represents the frequency of occurrence. The mean of the grey level intensity histogram is represented by the vertical red line. The mean of the positive pixels (MPP) is calculated from the grey level intensity values above zero (within the yellow box). This figure was adapted from Lubner et al. 2017[68].

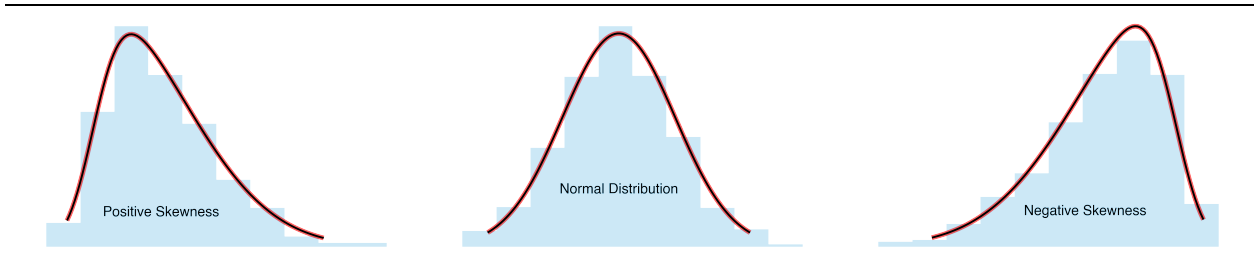


Figure 1.6 First Order Skewness as a Radiomic Feature

Applied to CT imaging, the x-axis would be CT attenuation value or Hounsfield units, and the y-axis would be the count/number of voxels. A positive skewness indicates that most values are clustered to the left whilst a negative skew indicates that most values are clustered to the right of the distribution[75]*. This figure was generated using the skewnorm function from the Scipy stats package in Python, and was adapted from Lubner et al. 2017*[47].

Miles et al. used computer modelling along with histopathological findings to determine what these TexRAD texture features mean in terms of image characteristics and their biological correlates[76]. Their findings are summarised in Table 1.2 and Table 1.3.

Table 1.2 Definitions of Histogram Parameters

Parameter	Definition	Corresponding image characteristics
Mean	The average value of the pixels within the region-of-interest	Changes approximately in proportion to the number of objects highlighted and their mean brightness (dark objects are negative)
Standard deviation (SD)	A measure of how much variation or dispersion exists from the average (mean value). A low SD indicates that the data points tend to be very close to the mean; high SD indicates that the data points are spread out over a large range of values.	Increases approximately in proportion to the square root of the number of objects highlighted and their mean intensity difference compared to background (i.e. dark and bright objects are both positive).
Skewness	A measure of the asymmetry of the histogram. The skewness value can be positive or negative. A zero value indicates that the values are evenly distributed on both sides of the mean.	Reflects the average brightness of highlighted objects (predominantly bright objects give positive values, predominantly dark objects negative values). Tends to zero with increasing number of objects highlighted. Moves away from zero with intensity variations in highlighted objects.
Kurtosis	A measure of the peakedness of the histogram. The kurtosis value can be positive or negative. A positive kurtosis indicates a histogram that is more peaked than a normal distribution. A negative kurtosis indicates that histogram is flatter.	Inversely related to the number of objects highlighted (whether bright or dark). Increased by intensity variations in highlighted objects.
Adapted from Miles et al.[76] *Cancer Imaging* **13,** 400–406 (2013).		

Table 1.3 Pathologic Correlates for CT texture measurements using TexRAD

Tumour type	Histological marker	Parameter (association)	Reference
NSCLC	Hypoxia: pimonidazole (extrinsic marker)	SD (+)	Ganeshan et al.[67]
Colorectal with KRAS mutation	Angiogenesis: CD105	Skewness (+)	Ganeshan et al.[77]
Colorectal without KRAS mutation	Angiogenesis: VEGF	SD (-)	Ganeshan et al.[77]
Glioma	Tumour grade	SD (+)	Skogen et al.[78]
Adapted from Miles et al.[76] *Cancer Imaging* **13,** 400–406 (2013).			

NSCLC, non-small cell lung cancer; SD, standard deviation; VEGF, vascular endothelial growth factor

Of the first order statistics derived by the filtration-histogram method, SD has most commonly been associated with specific histological markers as shown in Table 1.3. Ganeshan et al. found a positive association between SD and hypoxia in non-small cell lung cancer[67] and Skogen et al. found a positive association between SD and glioma tumour grade[78].

SD and the other texture features may be considered as an objective method to quantify the heterogeneity of an ROI. This lends itself well to quantifying tumour heterogeneity and discriminating between benign and malignant lesions since malignancy is associated with greater heterogeneity. Hu et al. extracted texture features from computed tomographic images of colonic polyps[79]. These texture parameters were able to differentiate between neoplastic colonic polyps and non-neoplastic polyps with an area under the receiver operating characteristic curve (AUC) of 0.80. Furthermore, CT texture features have been identified as independent predictors of survival for patients with oesophageal cancer, colorectal cancer and head and neck cancer[68].

A recent review summarized the findings of histopathological studies and their association with texture features[68]. Overall, TexRAD texture parameters are correlated with hypoxia, angiogenesis, genetic mutations and the different hallmarks of cancer[67].

1.3.2.2 PyRadiomics: open-source Python package for biomedical radiomics analysis

PyRadiomics is an open-source Python package, developed by van Griethuysen et al.[80], for the standardisation of radiomic feature extraction from medical images (including CT, PET and MRI images). This radiomics package enables the extraction of first order features, including those available in TexRAD, as well as higher-order radiomic features such as those derived from the grey level co-occurrence matrix or those derived using a run-length matrix which analyses spatial interrelationships between grey level values in a specific direction, see Figure 1.7.

PyRadiomics can be used with the freeware 3D slicer as the front-end graphical user interface (GUI) or with native Python in a Jupyter notebook or from the command line interface (CLI). The majority of the radiomic features calculated with PyRadiomics are in compliance with the radiomic feature definitions described in the Imaging Biomarker Standardisation Initiative (IBSI)[81] and these definitions as well as instructions on how to use this package is fully documented at https://pyradiomics.readthedocs.io/.

PyRadiomics provides functions for additional image pre-processing such as z-score normalisation, image resampling and filtering (including the LoG used in TexRAD). Therefore, radiomic features can be calculated on the original image or from the filtered images. There are several radiomic feature classes available in the package including: first order features, shape features, grey level co-occurrence matrix (GLCM) features, grey level size zone matrix (GLSZM) features, grey level run length matrix (GLRLM) features, neighbouring grey tone difference matrix (NGTDM) features and grey level dependence matrix (GLDM) features. These are defined in Table 1.4.

Table 1.4 PyRadiomics Texture Feature Classes

Abbreviation	Feature Class	Description
IH	First Order Intensity Histogram Statistics	Metrics are derived from the image histogram, similar to TexRAD, see Figure 1.5.
GLCM[82,83]	Grey Level Co-occurrence Matrix	Calculates the spatial dependency of grey levels in an image by expressing how combinations of the discretised grey levels of neighbouring pixels are distributed along different directions of the image.
GLRLM[84]	Grey Level Run Length Matrix	These matrices calculate the run length of a grey level in a certain direction, defined as the length in number of pixels, of consecutive pixels with the same grey level value.
GLSZM[85]	Grey Level Size Zone Matrix	Quantifies grey level zones in an image, defined as the number of connected voxels with the same grey level intensity.
GLDM[86]	Grey Level Dependence Matrix	These matrices quantify the grey level dependencies in an image, defined as the number of connected voxels within a certain distance that are dependent on the centre voxel.
NGTDM[87]	Neighbouring Grey Tone Difference Matrix	These matrices combine the sum of the grey level differences of voxels with a certain intensity value with the average discretised grey levels of a neighbourhood with a certain distance from the actual voxel.

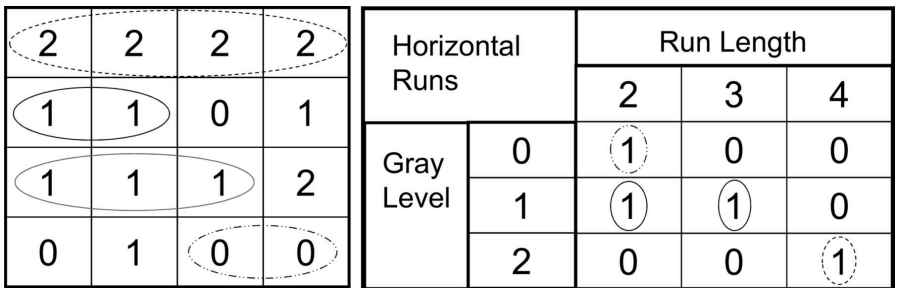


Figure 1.7 Grey level run length matrix

The GLRLM matrix enumerates the frequency of horizontal runs of a particular grey level value of a particular length. For example, the frequency of a horizontal run of 1s that are of length 2 is 1, subsequently the corresponding row in the GLRLM is populated accordingly. Reproduced from Buch et al. 2015[88]

There is greater flexibility in the number and nature of the radiomic features extracted with PyRadiomics compared with TexRAD. This leads to greater variation in the number and type of radiomic features extracted between different radiomic research studies that use PyRadiomics. There is also an increase in high-dimensional, low sample size data analyses for which feature selection or dimensionality reduction steps are typically required prior to building predictive models[89]. For example, Yang et al. used PyRadiomics to extract 1080 CT radiomic features from pleural nodules in NSCLC patients in order to identify patients with dry pleural dissemination (DPD)[90]. They used the least absolute shrinkage and selection operator (LASSO) binary regression classifier (which incorporates feature selection into its modelling and shrinks the coefficients of non-predictive features to zero) and identified a 'radiomics signature' composed of 10 radiomic features including First Order: Skewness, GLCM: Cluster Shade, GLCM: Maximum Probability, Wavelet GLCM: Cluster Shade and Wavelet: GLRLM: Long Run High Grey Level Emphasis that was statistically significantly different between the non-DPD and the DPD group ($p<0.001$)[90].

Using PyRadiomics to extract texture features from baseline CT scans of 207 patients with pancreatic ductal adenocarcinoma and incorporating this as input for a random forest machine learning classifier, Kaissis et al. were able to build a model that could predict the molecular subtype of the tumour: quasi-mesenchymal versus non-quasi-mesenchymal, with an AUC of 0.93 ± 0.01 on the independent test set[91]. This suggests that radiomic features, which are imaging-derived metrics, can potentially reflect the molecular phenotypes of tumours and reflect biologically relevant characteristics. For example, certain PyRadiomics CT radiomic features have been found to be positively correlated with disease progression in patients with gastrointestinal stromal tumours including GLCM: Inverse Difference ($p=0.012$, HR 3.83; 95% CI 1.697-8.611) and NGTDM: Coarseness ($p<0.001$, HR 3.156, 95% CI 1.554-6.411), whilst other radiomic features, particularly the second-order and higher-order PyRadiomic features have been associated with a number of tumour mutations[92]. As such, in the field of oncology, radiomic data is being combined with other '-omic' data, such as genomics to give rise to 'radiogenomics' which correlates tumour genotypes with image phenotypes in order to better understand the patient and the disease process on multiple levels[93]. Therefore, radiomics presents an opportunity to advance our progress in precision medicine[94].

Applications of PyRadiomics beyond oncology are emerging within the field of cardiology and stroke medicine. Liu et al. extracted 12 morphological (shape) features from 719 intracranial

aneurysms to predict aneurysm stability using machine learning models and found that Flatness was the most important radiomic predictor according to LASSO regression[95]. Another interesting application of PyRadiomics involved the extraction of a set of radiomic features to help quantify vascular collateral development from brain CT angiography images of 30 patients who experienced an ischaemic stroke. The images were filtered with a 'vesselness' filter prior to radiomic feature extraction and they found that out of the 75 radiomic features extracted, there were statistically significant differences between 57 of these between damaged and intact areas of the brain. Damaged areas were associated with lower GLCM: Contrast, Difference Variance and Joint Entropy compared with intact areas[96].

Currently, the most comprehensive radiomics analysis using PyRadiomics in cardiology relates to the development of a radiotranscriptomic signature from coronary CT angiography scans for coronary artery perivascular adipose tissue (PVAT)[97]. Using 3D slicer, the GUI that encapsulates PyRadiomics as an interactive plugin, 843 radiomic features including shape, first order and higher-order radiomic features were extracted from segmented coronary adipose tissue around the right and left coronary artery, resulting in 1686 radiomic features per patient. 82% of these features were identified as stable with an intraclass correlation coefficient (ICC) of 0.9 on inter-observer analysis, resulting in 1391 radiomic features that were subsequently fed into a random forest classifier with recursive elimination to create a PVAT radiomics signature. Gene expression studies were conducted in parallel looking at markers of vascularity (*CD31 expression)*, adipose tissue fibrosis (*COL1A1 expression)* and inflammation (*TNF expression)* and were found to correlate with the first order statistic: wavelet-transformed mean attenuation, and other higher-order radiomic features[97]. The prognostic value of the radiomics signature, referred to as the fat radiomic profile (FRP), was assessed in the prediction of major adverse cardiac events (MACE) on an external validation dataset (SCOT-HEART). The coronary FRP outperformed traditional cardiovascular risk factors, the coronary calcium score, coronary stenosis and the presence of high-risk plaque features for MACE prediction ([C-statistic] = 0.126, $P<0.001$)[97]. Therefore, this study demonstrates the potential for CT radiomic features to improve patient risk stratification and that radiomic features provide a potential non-invasive method to capture coronary inflammation and atherosclerotic disease processes.

1.3.3 Further Applications of Radiomics in Cardiology

The range of radiomics applications in cardiology is even greater when we consider radiomic studies not limited to TexRAD or PyRadiomics. Early work relating to the texture analyses of atherosclerosis were originally applied to ultrasound imaging[98,99]. Christodoulou et al. extracted first order and higher-order statistical texture features from 230 high-resolution ultrasound images of carotid plaques. These texture features were then fed into machine learning classifiers (self-organising map classifiers) to differentiate between symptomatic and asymptomatic carotid plaques, reaching an average AUC of 73.1%[98]. Awad et al. investigated different categories of texture features to include Fourier power spectrum and Laws texture energy features to assess the ability of 3D carotid ultrasound radiomic features in the classification of carotid arteries from patients taking atorvastatin versus placebo[99].

An interesting application for MRI imaging came from Larroza et al. in 2017[100–102]. They manually segmented infarcted and non-infarcted areas of the myocardium in the left ventricle of 10 patients. First- and second-order texture features were subsequently derived using open-source software (MaZda, Institute of Electronics, Technical University of Lodz, Poland) from these regions-of-interest. A machine learning approach was then used to identify the most discriminative features for identifying infarcted areas of the heart. The study revealed that first order features were the most important features for the classification of infarcted or non-infarcted regions and they could be used to provide an automatic infarcted myocardium segmentation solution.

In addition, radiomics analysis of MRI images with native T1 mapping has been applied to the differentiation of hypertrophic cardiomyopathy (HCM) from hypertensive heart disease (HHD). Neisius et al.[103] performed radiomics analysis on 232 subjects and identified six texture features with the best discriminatory capacity between HHD and HCM, and an internally validated accuracy of 80.0%. This proof-of-concept study demonstrates the potential for radiomics in cardiology to solve clinical problems, along with the need for further validation.

In recent years, radiomics analyses have extended to CT angiography, particularly focusing on coronary CTA. For example, the napkin-ring sign (NRS) is a subjective qualitative parameter derived from CT angiography images of the coronary arteries that is associated with major adverse cardiac events[59]. Radiomic features were extracted from CTA scans to predict the

presence of the NRS[104], which can often be difficult to identify reliably. In this study, the coronary plaques were manually segmented and 4440 radiomic parameters were extracted using a software tool called Radiomics Image Analysis, developed using the R programming environment. Using the permutation test of symmetry, Kolossváry et al. found 20.6% of the radiomic features present in NRS-positive plaques versus NRS-negative plaques were different and could be used for differentiation between the two types of plaque[105].

In another study, Kolossváry extracted 935 CTA radiomic features from 44 plaques of 25 patients and assessed their ability to identify several outcomes: 1) predict the presence of attenuated plaque as determined by IVUS, 2) predict the presence of thin-cap fibroatheroma as determined by OCT and 3) predict the presence of increased NaF uptake. Compared with conventional CT parameters, the radiomic features were better able to predict the outcomes e.g. the radiomic feature *fractal box counting dimension of high attenuation* vs. non-calcified plaque volume had an AUC of 0.72 (CI: 0.65-0.78) vs an AUC of 0.59 (CI: 0.57-0.62)[106]. The AUC values were determined by stratified five-fold cross-validation with 1000 repeats. Although independent validation would be required to verify these findings, this initial study suggests that CTA radiomic studies have potential prognostic value.

Nam et al. 2019[107] evaluated the use of cardiac CT radiomic features extracted from 3-dimensionally segmented ROIs of periprosthetic masses in the differentiation between causes of prosthetic valve obstruction (PVO) in patients who had undergone prosthetic valve replacement. This retrospective study included 39 periprosthetic masses in 34 patients and included 52 radiomic features (first order, size and shape features, second-order features from GLCM and GLRLM and wavelet features) calculated using in-house software, in a LASSO regression model. The AUC of the radiomic score for the identification of PVO cause, as determined by 10-fold cross validation was 0.876.

1.4 Machine Learning

Arthur Samuel informally defined 'machine learning' (ML) in 1959 as the field of study that gives computers the ability to learn without being explicitly programmed[108]. This was formally defined by Tom Mitchell in 1997 as "a computer program is set to learn from an experience E with respect to some class of tasks T and performance measure P if its performance on T, as measured by P, improves with experience E."[109] ML is a sub-section of the broader concept of artificial intelligence (AI) whose motivation is the creation of non-biological 'intelligence' and which lies at the intersection between computer science and statistics as shown in Figure 1.8.

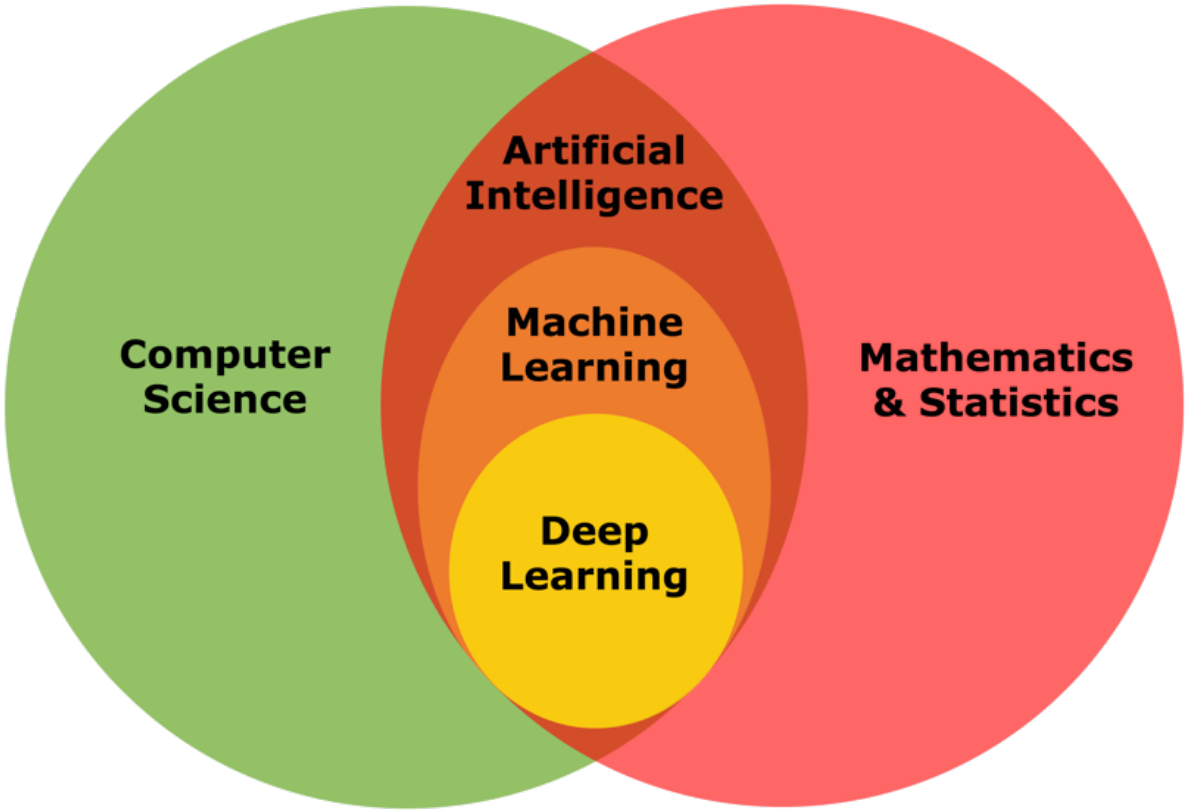


Figure 1.8 Intersection of Different and Related Fields

Machine learning is a sub-section of artificial intelligence (AI). These fields lie at the intersection between computer science, mathematics and statistics. Deep learning is a branch of machine learning and builds upon artificial neural networks.

Machine learning algorithms are suited to classification tasks such as email spam-filtering or medical image segmentation, regression problems such as predicting house prices or predicting patient systolic blood pressure[110] and clustering to find similar groups. Depending on the data available, there are different types of learning approaches as illustrated in Figure 1.9.

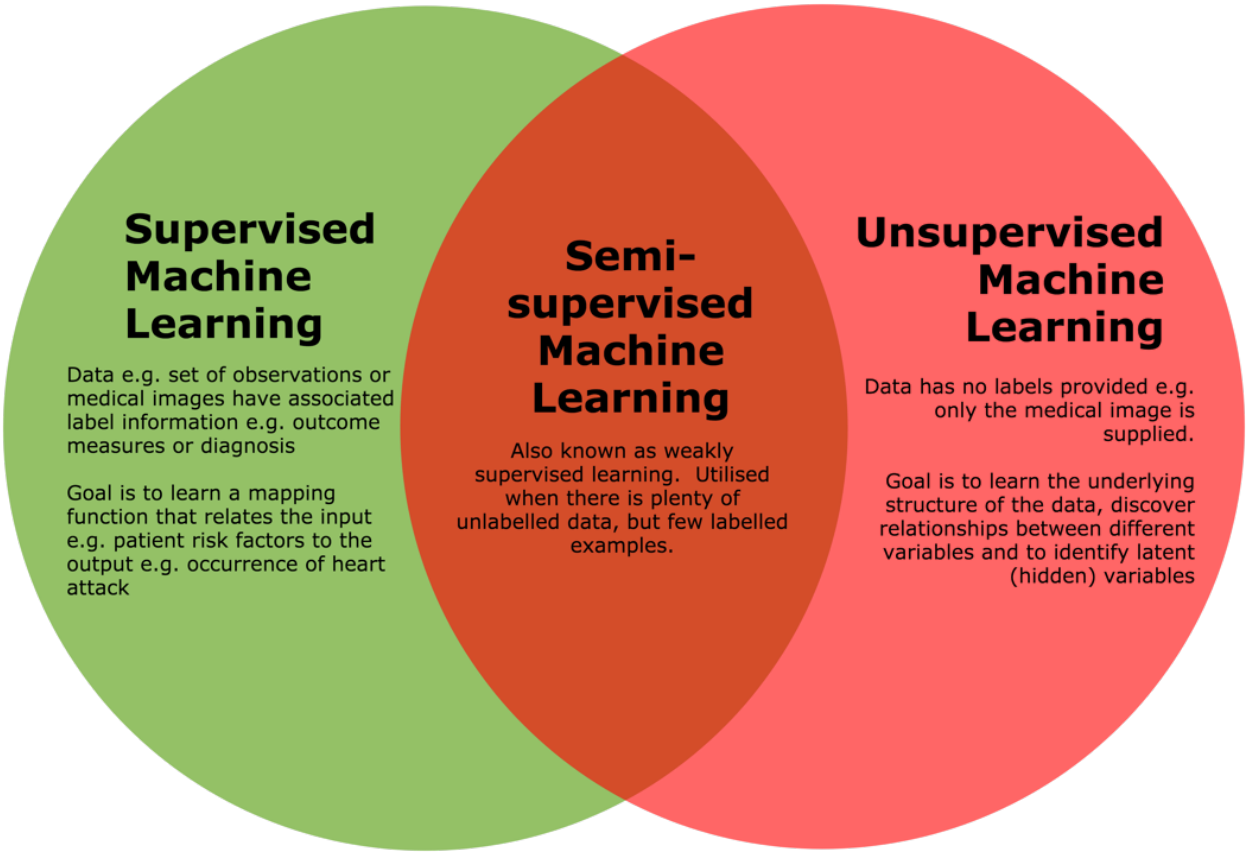


Figure 1.9 Machine Learning Approaches

Supervised machine learning requires training data with its associated labels in order to learn a mapping function from the input to the label. Once trained, test data can be provided for a predicted label output. Unsupervised machine learning does not require any labels and can discover different patterns in the data. Semi-supervised machine learning occurs when only some of the data has associated labels.

1.4.1 Supervised Machine Learning

The key characteristic of supervised ML is the ability to learn by training examples. During the training procedure, the training data and the associated ground truth labels for that data (sometimes referred to as 'targets') are both required[111]. This data-driven approach involves the learning of a mathematical model based on the observed 'training' data, essentially a mapping function y = f(x) from the input (x) to the output (y). The goal is to approximate the mapping function so well that for new values of x, we can predict y. Supervised learning is therefore employed for tasks such as classification and regression.

'Learning' or 'training' in machine learning usually refers to the iterative procedure of measuring the gap between current prediction (e.g. predicted label = symptomatic patient) against a ground truth (e.g. actual label = asymptomatic patient) using an evaluation metric

known as the 'objective function' (also known as the 'cost' or 'loss' function). The ultimate aim is that with further training, the model adapts to reduce this gap i.e. until the predicted label matches the actual label. The supervised machine learning workflow is shown in Figure 1.10.

Traditional ML algorithms require an explicit feature extraction step in the supervised machine learning workflow, known as feature-based machine learning. This means that we have to make active choices about what aspects of the data we feed into the ML algorithm which includes feature engineering and feature selection. For example, when using a traditional ML algorithm, we would not feed in the raw pixels of CT images, but *rather features derived from that image such as the coronary calcium score, or luminal stenosis*. Many thousands of features can be hand-crafted and derived from the CT images and feature selection involves choosing which features are most relevant for the chosen task.

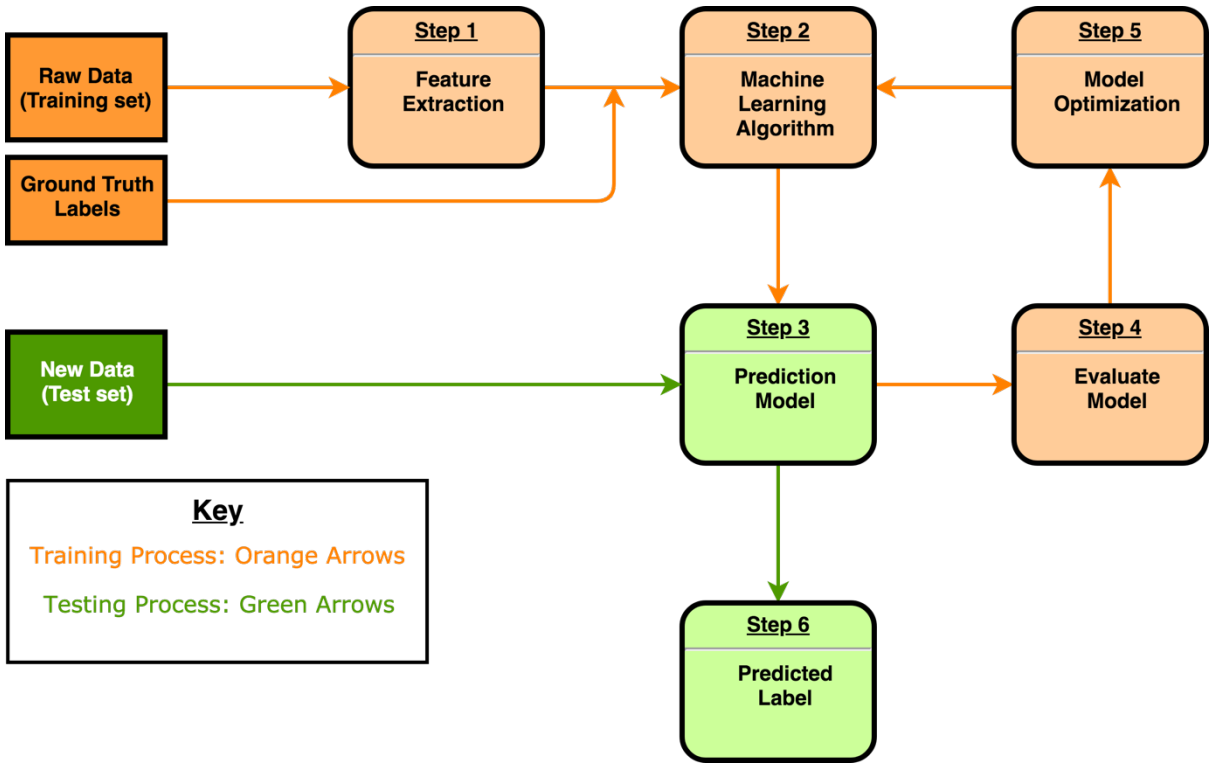


Figure 1.10 Supervised Machine Learning Workflow

During the training process, the training data and the associated ground truth labels are fed into the machine learning algorithm. The predictions of the ML model are evaluated against the ground truth and this is used to adjust the parameters of the ML model to optimize its performance. Once a satisfactory level of a performance metric e.g. accuracy has been achieved, the model's performance is evaluated on new data. This is known as the testing process.

The accuracy of these classification algorithms depends on the prior feature engineering process and so is constrained by the extent of prior expert knowledge. Nevertheless, classical ML algorithms have relatively low computational requirements as these models involve fewer parameters compared to deep learning algorithms. Following the principle of parsimony: "when you have two competing theories that make exactly the same predictions, the simpler one is better" (a variation of Occam's razor)[112], the simpler feature-based machine learning solution may be more appropriate than a deep learning solution if the models offer the same predictive ability. Furthermore, classical ML algorithms have shown success in several branches of medicine including incorporation into computer-aided detection systems for breast imaging[113,114], and oncological medical image segmentation[115] applications. A description of different supervised classical machine learning algorithms is provided in Table 1.5.

Table 1.5 Definitions of Classical Machine Learning Algorithms

Algorithm	Definition
Support Vector Machine (SVM)[116117]	Projects training points into a high-dimensional space and attempts to draw a separating line/hyperplane through the training points with the aim of maximizing the margin between different groups.
K-Nearest Neighbours (KNN)[118]	KNNs classify new data points based on a distance metric (similarity measurement) from old data points. The new data point is classified by the label of the majority of its neighbours.
Logistic Regression (LR)[119]	A modified version of linear regression for binary classification that produces a probability (value between 1 and 0) that the data belongs to a given class.
Decision Tree (DT)[120]	A conveniently interpretable ML algorithm that splits the data sequentially based on features that result in the largest information gain. A decision tree consists of a series of decisions, however, the DT may overfit on the dataset.
Random Forest (RF)[121]	An ensemble method built from many decision trees to provide better generalization.
Artificial Neural Network (ANN)[122]	Artificial neurons are arranged in layers with information being passed through the consecutive layers.

1.4.2 Unsupervised Machine Learning

In contrast to supervised learning, there is no predicted outcome in unsupervised learning and there is no corresponding ground truth label associated with the input data. Unsupervised learning can be useful in terms of discovering the underlying patterns of the data. It is often used for feature engineering to find features that can be subsequently used in a supervised model for disease prediction.

Classical unsupervised ML algorithms include clustering approaches (k-means clustering) and dimensionality reduction approaches such as principal component analysis. The dimension of a dataset refers to the number of features derived from that dataset. For example, feeding in a raw medical image of 512 x 512 for pixel-level analysis would result in 262,144 features for that one medical image alone. Dimensionality reduction algorithms can facilitate the identification of the most important components of the data, helping to improve the accuracy of predictive models and reduce the required training time.

There are more sophisticated deep learning approaches, including autoencoders[123], which are able to discover different representations of the input data. An autoencoder encodes the input into a highly compressed latent space and a decoder function attempts to reconstruct the original input from this constrained compressed representation. Unsupervised representation learning is often used as a pre-training step before a supervised machine learning approach is employed in order to reduce necessary training times and improve the performance of predictive models.

1.4.3 Machine Learning: Cardiovascular Applications

One of the first studies to embrace 'Big Data' and machine learning for cardiovascular risk prediction, Weng et al.[124], applied four different machine learning algorithms: random forest, logistic regression, gradient boosting machines and neural networks to a large dataset of routine clinical data relating to over 370, 000 patients from the UK general primary care population. The dataset originated from a prospective study that followed participants initially free from cardiovascular disease. At follow-up, close to 25, 000 cardiovascular events (6.6%) occurred. The performance of the ML algorithms was compared with the ACC/AHA risk prediction algorithm[125] in their ability to identify individuals who would develop CVD and those that would not.

In ML terms, this was a binary classification task in which each ML algorithm was trained on 78% of the dataset (n=295, 267). During training, ML algorithms automatically adjust their parameters to learn which clinical variables are important in predicting the outcomes: (1) cardiovascular event occurring or (2) not. The remaining 22% of the data (n=82, 989) was kept separate from the training step, and was used to assess the predictive accuracy of the models i.e. the patient features are fed into the trained ML model which outputs a predicted label for

patient outcome. The headline result was that all ML algorithms improved prediction accuracy, assessed by the AUC, over the ACC/AHA risk model: from AUC 0.728 to 0.745 with random forest, 0.760 with logistic regression, 0.761 with gradient boosting and the best performance came from neural networks at AUC 0.764.

This study indicates that machine learning approaches do have the potential to improve cardiovascular risk prediction, yet there is still plenty of room for improvement. The predictive accuracy of the models also depends heavily on the choice of ML model and the features i.e. the clinical variables that are supplied to the model. Current, non-ML risk prediction algorithms implicitly assume a linear relationship between risk factors and outcomes. However, these results suggest that there are non-linear, complex interactions between risk factors that remain to be explored and exploited using machine learning[84].

Ambale-Venkatesh et al.[126] evaluated the performance of random forests in predicting different cardiovascular outcomes using data from the Multi-Ethnic Study of Atherosclerosis cohort (MESA). This comprised 6814 participants initially free from CVD that were followed up for over 12 years. The unique approach of this study was the incorporation of 735 clinical variables ranging from imaging to blood tests and questionnaire responses as the input features. The top 20 predictors for each cardiovascular outcome, identified by the model, were insightful. For example, imaging, electrocardiography (ECG) and serum biomarkers emerged as strong predictors as opposed to the traditional cardiovascular risk factors. In terms of evaluating predictive accuracy, this study did not use AUC, but instead they used the Brier score which is the mean squared error between the predicted probability of the outcome and the actual outcome. This highlights one of the limitations in evaluating ML models across different studies and comparing model performances - there is currently a lack of standardisation and best practice guidelines until recently[127–129], which is one of the barriers to the widespread adoption of ML into clinical practice.

Motwani et al.[130] derived 44 coronary computed tomographic angiography metrics such as number of segments with non-calcified, mixed or calcified plaques and combined them with 25 clinical variables such as age, gender and hypertension from a multicentre prospective registry for machine learning analysis. They found that these non-invasive imaging-derived metrics were able to significantly improve all-cause mortality risk prediction accuracy in patients with suspected CAD, particularly when incorporated with clinical non-imaging data.

This suggests that imaging data, even simple image-derived metrics (rather than the raw voxels), contain additional predictive value that might improve both risk prediction and stratification.

1.4.4 Machine Learning Considerations

1.4.4.1 Overfitting and Underfitting

Overfitting refers to the scenario whereby the machine learning model learns to perform well on the training data, without the ability to perform well on other datasets i.e. generalise. This can happen when the model is more flexible and complex than is necessary and mistakes the noise within the dataset for a signal that represents the mapping function between the input data and the output label[131]. When the training data is small and limited, overfitting becomes more likely. Under-fitting is the opposite of this problem, in which the machine learning model that is used is not flexible enough or is too simple to capture the relationship between the input data and the output label.

1.4.4.2 Model Validation

In order to evaluate the effectiveness of the mapping function learnt by the machine learning model between the input data and the output labels and to estimate how well the model would perform on a dataset that is not the training dataset, there are several possible methods: a) train-test split, b) cross-validation, c) bootstrapping and d) external validation.

A) Train-test split

The training set refers to the set of data that is used to train the ML model. The test set contains data that the ML model was never trained on. This is an independent dataset and is the 'new' data to test the predictive accuracy of the trained ML model. If the dataset is large enough, a validation set[132], also known as the tuning set, may be available. This is distinct from the test set and the validation set is part of the data that is fed to the ML model during the training process that is used to find out what are the optimal hyperparameters for the model, rather than to update the model parameters. Hyperparameters are separate to the model parameters. For example, in a neural network, the model parameters refer to the weights and biases. During training, the neural network uses the data in the training set to update these weights and biases

in order to minimise the error between the model's predictions and the ground truth labels. Therefore, ML model parameters are automatically tuned by the ML algorithm during training and is a data-driven process. However, hyperparameters are chosen a priori by the human and can impact the model's learning process. Examples of hyperparameters include: the number of epochs the model is trained for i.e. the number of times the model is permitted to go through the training dataset, the number of layers within a neural network and the learning rate which controls the magnitude of how much the weights and biases are adjusted during each update. It is not always clear what the optimal values for the hyperparameters are, and so a selection of possible values is often provided to try out on the validation set to figure this out. The validation dataset may also be used to monitor whether a ML model is overfitting on the training dataset by monitoring the learning curves of the ML model, see Figure 1.11.

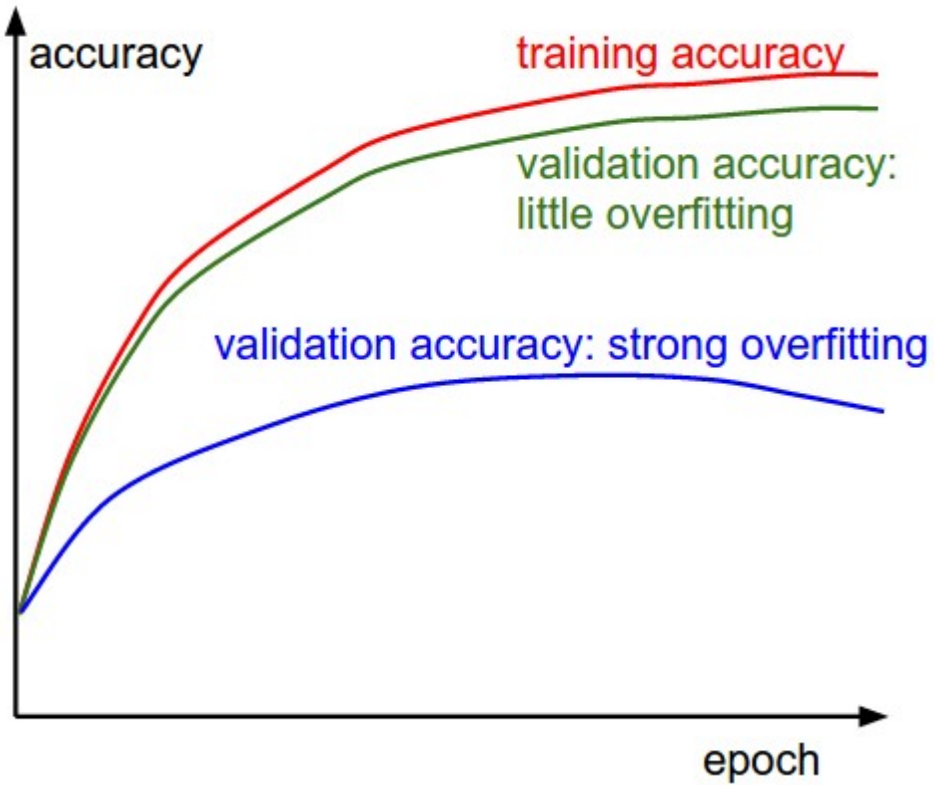


Figure 1.11 Learning curves to monitor ML model performance

The blue curve represents a validation accuracy curve that would be seen in the case of strong overfitting on the training set. The red curve plots the training accuracy against the number of epochs (the number of times the model has been through the training dataset). The training accuracy continues to increase, whilst the accuracy on the validation dataset does not continue to improve but starts to decrease in the blue curve. Reproduced from CS231n Convolutional Neural Networks for Visual Recognition website[133].

B) Cross-validation

Cross-validation involves splitting the available dataset into multiple training and test sets. For example, in 5-fold cross-validation, the data is split into 5 folds and each fold is split into a training set and a test set. In each fold, a machine learning model is trained on the training set and tested on the test set. The average performance of the models after the 5-folds is calculated to give an estimate of the ML model's performance on an external dataset, see Figure 1.12. Essentially, a different model is learnt in each fold during cross-validation. Typically, the machine learning model is retrained on all of the data before evaluating the performance on an independent dataset.

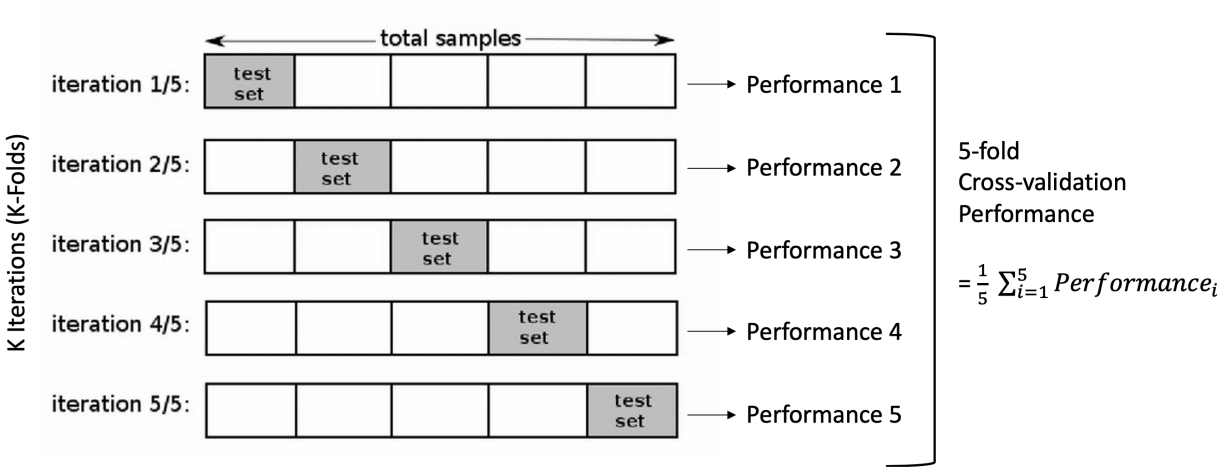


Figure 1.12 5-fold Cross Validation

This procedure results in 5 models fitted to different subsets of the data and tested on different parts of the data. The cross-validation performance is the mean of the performance estimates on each of the test sets in the 5 folds. Adapted from Comparison of Statistical Methods for Genomic Signature Extraction (2013) Chlis[134].

C) Bootstrapping

Cross-validation is referred to as a resampling technique, another such technique is known as bootstrapping. Bootstrapping involves repeatedly sampling from the dataset with replacement to produce estimates of statistics such as the machine learning model's accuracy[135]. Both cross-validation and bootstrapping are internal validation techniques that generate estimates of the model's ability to generalise on new datasets, without testing the model on an independent dataset.

D) External validation

In external validation, the machine learning model's performance is evaluated on an independent dataset that has similar properties to the dataset that the model was developed on but may come from a different hospital or scanner, for example. This gives the most reliable measure of model generalisability compared with the other techniques.

1.5 Deep Learning

Deep learning (DL) is a state-of-the-art machine learning technique that builds upon deep neural networks. Neural networks (NNs) were inspired by the functionality of the brain in terms of parallel processing and distributed representations and although they are an oversimplification of human neural processing, they form the building blocks of various DL architectures and have been very successful in complicated tasks such as image recognition[114,132,136] and natural language processing.

An individual neuron is the simplest unit of a neural network and its structure is illustrated in Figure 1.13. Each neuron consists of only two relatively simple operations that it applies to the input: (1) a linear (affine) transformation which is similar to linear regression (2) an activation function which helps to model non-linear functions[137]. The weights (w) and biases (b) shown in Figure 1.13 are the parameters of the network in which the weights are the coefficients for the importance of each feature and a bias sets the baseline activation threshold for a neuron. The weights and biases are slowly updated through backpropagation in order to minimize its error. Backpropagation is a method for the iterative adjustment of the weights and biases, whereby it calculates the gradient via an optimization algorithm, such as stochastic gradient descent, that relates weights to error and works backwards through the model to update the parameters[137].

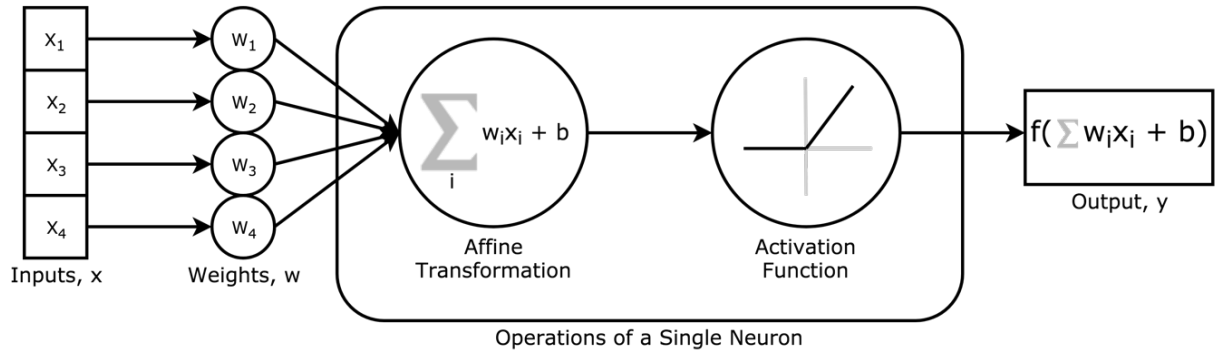


Figure 1.13 Structure of a Single Neuron in a Neural Network

The operations of a single neuron include a linear transformation of the input data, followed by an activation function which helps to model non-linear functions. The parameters of the neuron that can be adjusted during training are the weights associated with the inputs and the biases.

Groups of neurons that perform similar functions are aggregated into layers[138], see Figure 1.14. A neural network consists of consecutive layers, comprising an input layer e.g. raw medical image pixels, hidden layer(s), and an output layer e.g. outputs the prediction of the model: 'symptomatic'/'asymptomatic' label.

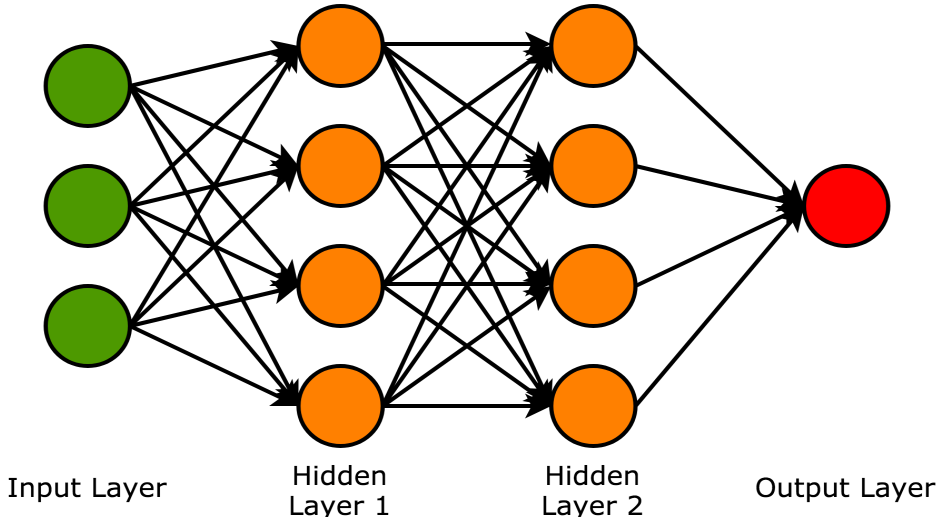


Figure 1.14 Structure of a Neural Network

A neural network consists of many neurons that are arranged into different layers. There is typically an input layer, hidden layer(s), and an output layer. The input layer has as many nodes as the number of features, the values are then multiplied by weights (represented by lines connecting the input nodes to the next layer of nodes). These nodes receive the products of the prior layer nodes/weights, sum them up and then apply an activation function to that sum, which determines the output value for this node. This continues for each layer until the final layer (the output layer) where the final decision is made. The layers between the input and output layers are referred to as 'hidden' layers. Traditional neural networks had 1-2 hidden layers, while current 'deep' neural networks have 10s to 100s of layers. Reproduced from Le et al. 2019[114]

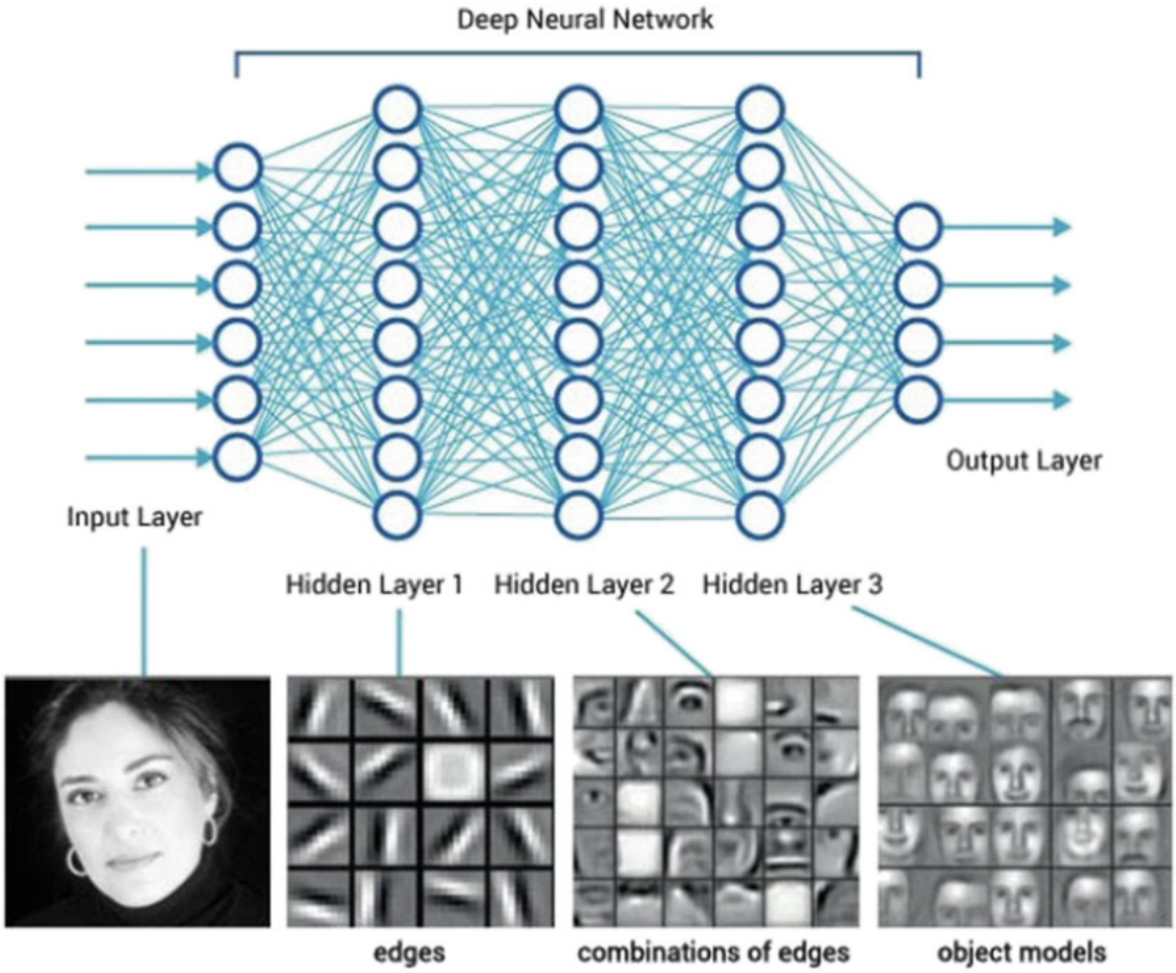


Figure 1.15 Structure of a Deep Neural Network

A neural network consists of many neurons that are arranged into different layers. There is typically an input layer, hidden layer(s), and an output layer. In the first layer, simple features are detected such as edges and lines. In the next layer this could be noses and mouths. The last layer of the network in a deep neural network is the classifier which gives the probability for the classes. Reproduced from B.M. ter Haar Romeny 2019[139].

The architecture of a neural network can be depicted as a directed computational graph, as in Figure 1.14 and Figure 1.15. In graph notation, each node (circle) is an operation and each edge (unidirectional arrow) indicates the flow of data through the graph. Even though a single neuron is very simple in its working, neural networks illustrate the concept of 'simplexity'[140] because the neurons in the network can be connected in several different and complex ways. The more complex the graph, the more relationships the DL algorithm can 'learn'. Therefore, we can think of a deep neural network model as a sequence of mathematical operations applied to the input to get an output and there can be millions of parameters (weights) in this mathematical function.

DL algorithms have the potential to reverse engineer any function as suggested by the Universal Approximation Theorem[141]: "any function can be arbitrarily approximated by a feedforward network with only one hidden layer." Please note that the theorem does not specify any time frame, nor does it bound the size of the hidden layer.

1.5.1 Convolutional Neural Networks

When dealing with computer vision tasks however, standard feedforward NNs (as depicted in Figure 1.14) are not the best structures to deal with images as they have no concept of spatial relationships and proximity related to the features of an image. The standard neural network converts an image into a linear vector as the input and looks at the pixels individually rather than looking at them together. This is fundamentally different to how humans visually process images: we group pixels together and see edges and lines, rather than focusing on individual pixels.

Convolutional Neural Networks (CNNs) are specialised deep learning structures characterised by their convolutional layers[138]. The convolution operation (also known as a kernel) acts as a filter on the pixel image matrix to extract spatially correlated features of the input image, similar to the receptive field of our photoreceptors. Convolution filters move across the input image data to create feature maps, with different convolutions resulting in different features being extracted from the image. As with other deep learning architectures, CNNs are composed of several layers which enables hierarchical learning. The earlier layers are considered to act similarly to the simple cells of the human primary visual cortex that learn low-level features such as edges in particular orientations in the image. Higher levels of abstraction, such as edges → motifs → parts → object, are a result of information being propagated through multiple stacked layers[142]. The last layer of the network architecture takes in these features for prediction or classification.

Whereas traditional ML, as described in Section 1.4.1, requires determination and calculation of features[143] from which the algorithm learns, deep learning automatically learns the important features as well as the weighting of those features to make predictions for new data. This is referred to as end-to-end training. Consequently, DL is more time consuming in terms of training as there can be millions of parameters to tune and hence DL typically requires high-

performance computing hardware. Nonetheless, DL approaches have their advantages including the ability to deal with increased complexity of input data (raw imaging data rather than statistically-derived features such as entropy and kurtosis) and compared to classical ML approaches, the DL models are able to learn much more interesting non-linear relationships between the input and the output.

One of the most famous CNNs is called AlexNet[144], so named after one of its designers Alex Krizhevsky. The network successfully won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[145] in 2012 and beat the other competing solutions by a margin of greater than 10%. The labelled training dataset for the ILSVRC consisted of over one million images of everyday objects and the task was to construct a model that can correctly classify the objects[144].

Medical image datasets that are available for research purposes, particularly well-curated and labelled datasets, are valuable and scarce resources. Several thousand images are considered a large dataset in the medical community, but this would be considered a relatively 'small' dataset in the wider deep learning community. Nonetheless, there are methods with which we can leverage the benefits of famous CNN architectures without the need for so much training data. One useful technique is known as transfer learning.

1.5.2 Transfer Learning

Transfer learning (TL) refers to the process of applying a model trained on a different dataset (such as natural images of everyday objects) to a different domain (medical images). The underlying idea is that the lower-level features that are learnt are useful in both natural images and medical image applications. With TL, we are able to leverage the power of sophisticated DL architectures, such as those designed by Google and Microsoft and adapt them for our own objectives. It is important to recognise however, the differences between the low-resolution, two-dimensional, colour (three-channel) images i.e. the natural images in ImageNet and medical images (higher resolution, multi-slice CT scans for example). Nevertheless, transfer learning has been successfully applied to several medical tasks such as the diagnosis of different skin lesions[136] and diabetic retinopathy[146] from medical images, reaching predictive accuracies that are on-par with human experts in certain settings.

1.5.3 Deep Learning: Cardiovascular Applications

Researchers at Google were able to use deep learning techniques for the prediction of cardiovascular risk factors and MACE within five years from retinal fundus images alone. Poplin et al. 2018 demonstrated that retinal fundus images contained information that were markers of cardiovascular disease and future cardiovascular risk[132]. They used the Inception-v3 neural network architecture (transfer learning) to predict the labels: gender (AUC=0.97), smoking status (AUC=0.71), major adverse cardiac events (AUC=0.70) and continuous variables including systolic blood pressure (mean absolute error within 11.23 mmHg).

The DL models were trained on data from the UK Biobank and EyePACs, totalling 284,335 patients and were tested on two independent datasets of 12,026 and 999 patients. Comparatively this is a much larger dataset than would be available in local research departments, and yet the Google researchers highlight that a limitation of this study was the relatively 'small' size of the dataset for deep learning. The predictive accuracy of the DL model for MACE was slightly sub-par compared with contemporary risk prediction approaches. This could be attributed to the fact that MACE outcome data was relatively rare and was only available for the UK Biobank data set (only 631 events occurred within 5 years of retinal imaging). Such an imbalanced dataset (where the number of data points for the minority class i.e. MACE << majority class i.e. no MACE) hindered the performance of the DL model.

The interest and application of deep learning to cardiovascular imaging has grown rapidly. Between January 1, 2017 to January 2019, there were over 80 original research papers regarding the application of deep learning to cardiovascular imaging according to a review by Litjens et al[147]. Use cases included image segmentation, calcification detection and obstructive vascular disease prediction. Figure 1.16 summarises some of the deep learning applications.

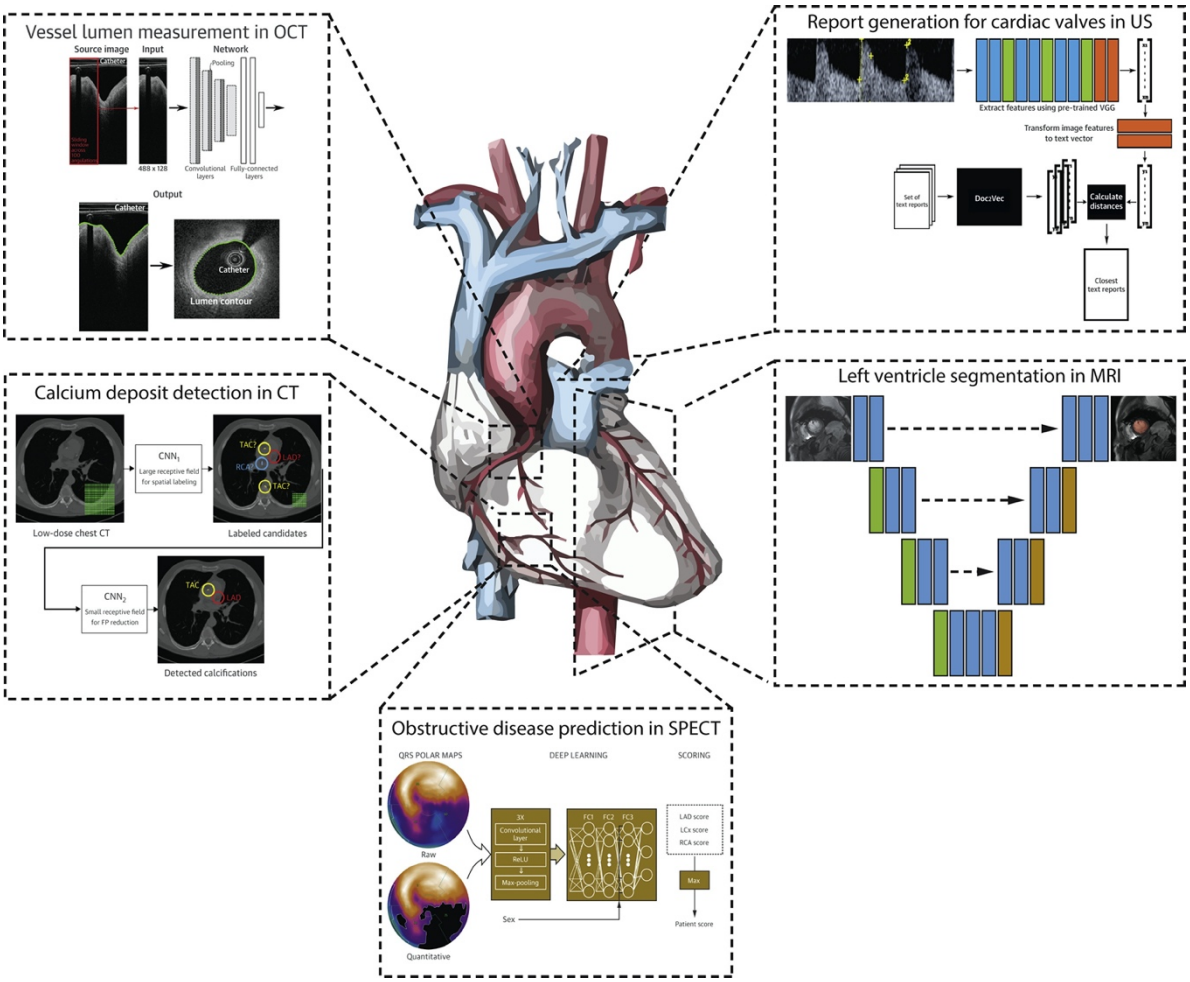


Figure 1.16 Applications of deep learning in cardiovascular imaging

Deep learning has been applied to different imaging modalities and uses cases including vessel lumen measurement, left ventricle segmentation, calcification deposit and obstructive disease prediction. CMR, cardiac magnetic resonance; CNN, convolutional neural network; CT, computed tomography; LCx, left circumflex; LAD, left anterior descending; OCT, optical coherence tomography; RCA, right coronary artery; SPECT, single-photon emission computed tomography; TAC, thoracic aortic calcification; US, ultrasound. Reproduced from Litjens et al. 2019 State-of-the-Art Deep Learning in Cardiovascular Image Analysis[147].

The take-home message from this review was that the applications of deep learning in cardiovascular imaging are diverse, with the majority of deep learning studies focusing on automating tasks such as calcium detection for calcium scoring, or image segmentation. The application of deep learning for prognostication is more limited as the underlying task is more difficult. The assumption is that there is a detectable signal for predicting patient risk from the medical images which may or may not be the case.

An additional challenge is that neural networks upon which deep learning solutions are based, can be "statistically impressive, but individually unreliable" and "can make mistakes that humans would not"[148]. Therefore, deep learning can help unravel new knowledge from medical images in ways that we would not expect and provides many opportunities for improving cardiovascular risk prediction on an individual and population-level. However, it is also imperative that we proceed with optimistic caution and assess these models as vigorously as any other potential clinical biomarker.

1.6 Aims and Hypotheses

The aims of this thesis were to evaluate the potential application of radiomics and machine learning for cardiovascular risk prediction. The specific objectives were:

- To assess the discriminatory ability of radiomic features for culprit versus non-culprit carotid arteries in cerebrovascular events and asymptomatic carotid arteries.
- To investigate the robustness of radiomic features to variations in the radiomics workflow.
- To investigate the association between radiomic features with functional PET imaging metrics, such as markers of inflammation.
- To evaluate the potential of feature-based machine learning with CT radiomics versus deep learning in multi-class classification of carotid types.

Main hypotheses:

- Culprit carotid artery lesions will display more CT signal heterogeneity than non-culprit lesions and asymptomatic carotid arteries.
- There will be differences in radiomic features between symptomatic individuals and asymptomatic individuals.
- There will be robust radiomic features for the prediction of culprit versus non-culprit carotid arteries.
- Radiomic features can be used as imaging biomarkers with the assumption that they will reflect certain underlying pathophysiological processes of atherosclerosis that will correlate with histological findings.
- Deep learning models will perform better than feature-based machine learning models.

1.7 Thesis Outline

Chapters 2 and 3 - focus on the application of radiomics to CT angiography and unenhanced CT scans of the carotid artery in single-slice and multi-slice analysis. We assessed the feasibility of deriving carotid artery imaging biomarkers for the differentiation between culprit (CC) and non-culprit (NC) lesions in symptomatic individuals (patients who have had a prior transient ischaemic attack or stroke), and then subsequently extended this to a multi-class problem to include differentiation of asymptomatic carotid arteries from asymptomatic individuals. We statistically compared the radiomic features of the different carotid artery types: asymptomatic, culprit and non-culprit, at first focusing on first order features derived from TexRAD, and then extending to higher-order statistical features with PyRadiomics. There were statistically significant differences in first order texture features between asymptomatic carotids and symptomatic CC or NC carotid arteries. There were differences between CC versus NC carotid arteries too. The differences between asymptomatic and symptomatic carotid arteries were greater than the differences between CC vs NC carotid arteries. Multi-slice analysis was more effective than single-slice analysis to reveal differences between carotid artery types using radiomics.

Chapter 4 - focuses on assessing the robustness of these radiomic features to variations in the radiomics workflow, with a particular focus on variabilities within image segmentation. Computer vision morphological operations were applied to regions-of-interests drawn around the carotid artery to mimic human under- and over-segmentation, reflective of actual intra- and inter-observer variability when ROIs are manually delineated. This was conducted under several different image processing settings to include pre-processing the images by resegmentation or normalisation, different resampling methods and discretisation methods. A set of robust and non-redundant radiomic features for the differentiation of culprit and non-culprit carotid arteries was identified for multi-slice and single-slice analysis, with and without resegmentation.

Chapter 5 – focuses on the application of machine learning for the identification of the culprit carotid artery. The predictive ability of the carotid calcium score, PET information and radiomic features both alone and together were investigated in multi-class classification of asymptomatic versus culprit versus non-culprit carotid artery status. To assess the biological

relevance of these radiomic features, we correlated the radiomic features with functional data (degree of inflammation as determined by PET imaging e.g. FDG mean of TBRmax), and histological findings from carotid endarterectomy specimens. Feature-based machine learning with radiomic features performed better than carotid calcification or PET imaging data as predictors alone.

Chapter 6 - Deep learning discovers visual features including both low-level and high-level features without human input and so unlike radiomics, deep learning features are not limited by expert knowledge. This chapter focuses on the application of deep learning to the carotid imaging dataset, investigating the use of transfer learning to overcome the limited dataset, and the use of 3D convolutional neural networks (CNN) for multi-class classification. We used the Grad-CAM method to visualise and interpret what the CNNs were looking at in the images to make their classifications decisions and investigated the robustness of deep learning decisions to perturbations of the original image. Although deep learning is a viable method to learn new features from the imaging data, the risk of overfitting is greater than with feature-based machine learning as the model's complexity and flexibility is much greater.

Chapter 7 - presents a summary of the PhD, evaluates the projects and highlights the challenges and opportunities for implementation into the clinical workflow. Finally, we explore potential future directions and opportunities for future research.

Chapter 2: Texture Analysis of Carotid CT and CT Angiograms with TexRAD

Chapter summary:

This chapter provides details of the materials and methods, results and a discussion in relation to the comparative analysis of first order radiomic features extracted using TexRAD in single-slice and multi-slice analysis of carotid CT scans (unenhanced and contrast-enhanced) for the identification of culprit, non-culprit and asymptomatic carotid arteries.

2.1 Background

A CT scanner typically consists of an X-ray source and a detector that rotates around the patient. CT images are reconstructed based on the attenuation of the X-ray photons as they pass through different tissues within the body which have different densities. The grey level values within a CT image represent Hounsfield units (HU), which are based on the attenuation values from the tissues relative to that of distilled water (arbitrarily designated 0 HU). Different tissue types correspond to different Hounsfield units, the approximate values are listed in Table 2.6. To better visualise the vasculature, iodine-based contrast media can be administered in CT angiography[149] (CTA).

Table 2.6 Different tissues and their associated Hounsfield unit ranges

Tissue	Hounsfield unit Ranges (HU)
Bone	+400 → +1000
Soft tissue	+40 → +80
Water	0
Fat	-60 → -100
Air	-1000

HU, Hounsfield units. Adapted from Hayashi et al. 2016[150]

2.1.1 Aims and Objectives

This chapter focuses on the texture analysis of unenhanced CT and CT angiography images of carotid vasculature in asymptomatic and symptomatic (stroke or transient ischaemic attack [TIA]) patients. The specific objectives for this chapter are as follows:

Primary objectives:

- To investigate whether there are statistical differences between first order radiomic features derived from CT angiography and unenhanced CT images of culprit and non-culprit carotid arteries in symptomatic patients and asymptomatic carotid arteries in asymptomatic patients (no prior cerebrovascular events).
- To evaluate differences in texture features derived from CTA images and those from unenhanced CT scans.

Secondary objectives:

- To evaluate the value of multi-slice analysis versus single-slice analysis in carotid artery radiomics.
- To stratify radiomic analysis of the carotid arteries by carotid plaque type and carotid stenosis severity.
- To evaluate the differences in the degree of carotid calcification in asymptomatic versus culprit versus non-culprit carotid arteries and its spatial distribution.

2.1.2 Hypotheses

- Culprit carotid artery lesions display more heterogeneity than non-culprit lesions in symptomatic patients.
- Symptomatic carotid arteries from symptomatic patients display more heterogeneity than asymptomatic carotid arteries from asymptomatic patients.
- There are statistical differences between the radiomic profiles of asymptomatic, culprit and non-culprit carotid arteries respectively.
- Multi-slice analysis is better than single-slice analysis to reveal radiomic differences.
- Non-culprit carotid arteries are more calcified than culprit carotid arteries.

2.2 Materials and Methods

2.2.1 Carotid CT Datasets

This retrospective study analysed unenhanced CT and CTA images pooled together from three previous observational research vascular imaging datasets (ICARUSS[151], VISION[152] and CHAI[153]) originating from a single institution (Addenbrooke's Hospital, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge, UK) between 2011 and 2016. All the series had appropriate ethical approvals in place and had similar inclusion and exclusion criteria which are listed in the published papers[151–153]. The different vascular imaging research studies had standardised imaging protocols which are detailed in section 2.2.1.1.

Kind permission was provided to use the anonymised carotid unenhanced CT and CTA images as well as related clinical information of the ICARUSS study from Dr Nicholas Evans (NE), Clinical Lecturer in Geriatric and Stroke Medicine; of the VISION study from Dr Jason Tarkin (JT), Clinical Research Fellow and Cardiology trainee; and of the CHAI study from Dr Francis Joshi (FJ), Cardiology Consultant.

Symptomatic patients had confirmed carotid artery-related ischaemic stroke or TIA within 3 months prior to imaging. Patients with haemorrhagic stroke, atrial fibrillation and/or complete vessel occlusions were not included in this study. In symptomatic patients, the carotid artery associated with precipitating the cerebrovascular event was deemed the 'culprit' carotid artery and was determined by its consistency with the clinical presentation of stroke (or TIA) symptoms, whilst the contralateral carotid artery was deemed the 'non-culprit' carotid artery. Endarterectomised carotid arteries that took place prior to imaging were excluded from analysis on the basis that any recurrent lesions may not be representative of atherosclerosis.

Patients classified as 'asymptomatic' had no prior TIA or stroke and came from one of three possible categories: (1) asymptomatic but with evidence of carotid atheroma (n=7 patients), (2) patients with stable angina (n=9 patients) or (3) patients that had experienced an acute coronary syndrome (ACS) event within the 3 months prior to imaging (n=9 patients). Table 2.7 demonstrates that the asymptomatic patients in this study came from the VISION and CHAI studies.

Table 2.7 Summary of Pooled Datasets

	ICARUSS	VISION	CHAI	Total
Patients	22	34	10	66
Symptomatic	22	13	6	41
Stroke	22	3	5	30
TIA	0	10	1	11
Asymptomatic	0	21	4	25
Carotids	44	68	20	132
Symptomatic	44	26	12	82
Culprit	22	13	6	41
Non-culprit	22	13	6	41
Asymptomatic	0	42	8	50

TIA, transient ischaemic attack

Figures 2.17-2.19 present flowcharts for the different vascular imaging datasets included in this study with a breakdown of the number of symptomatic and asymptomatic individuals, as well as culprit and non-culprit carotid arteries.

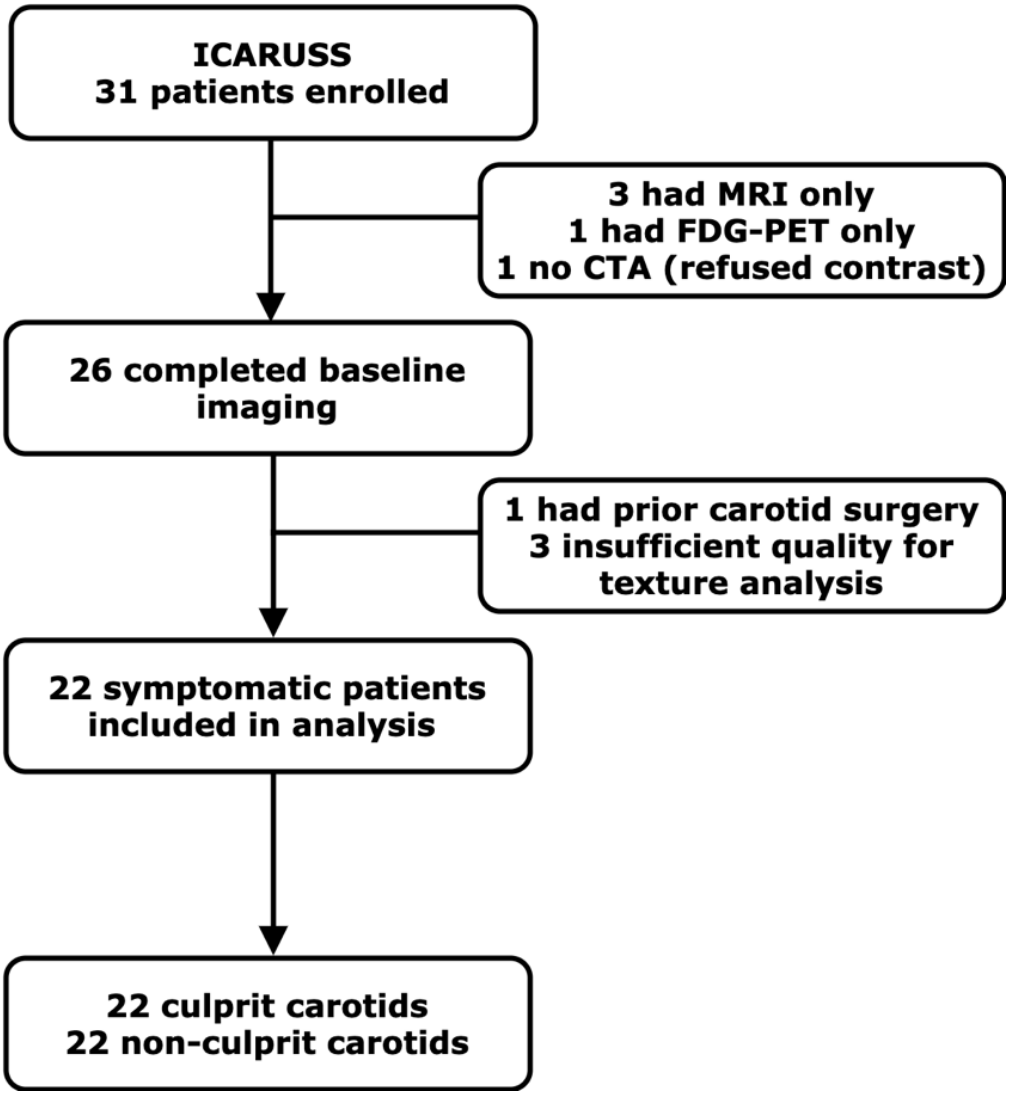


Figure 2.17 ICARUSS: study ran from 2014 to 2016

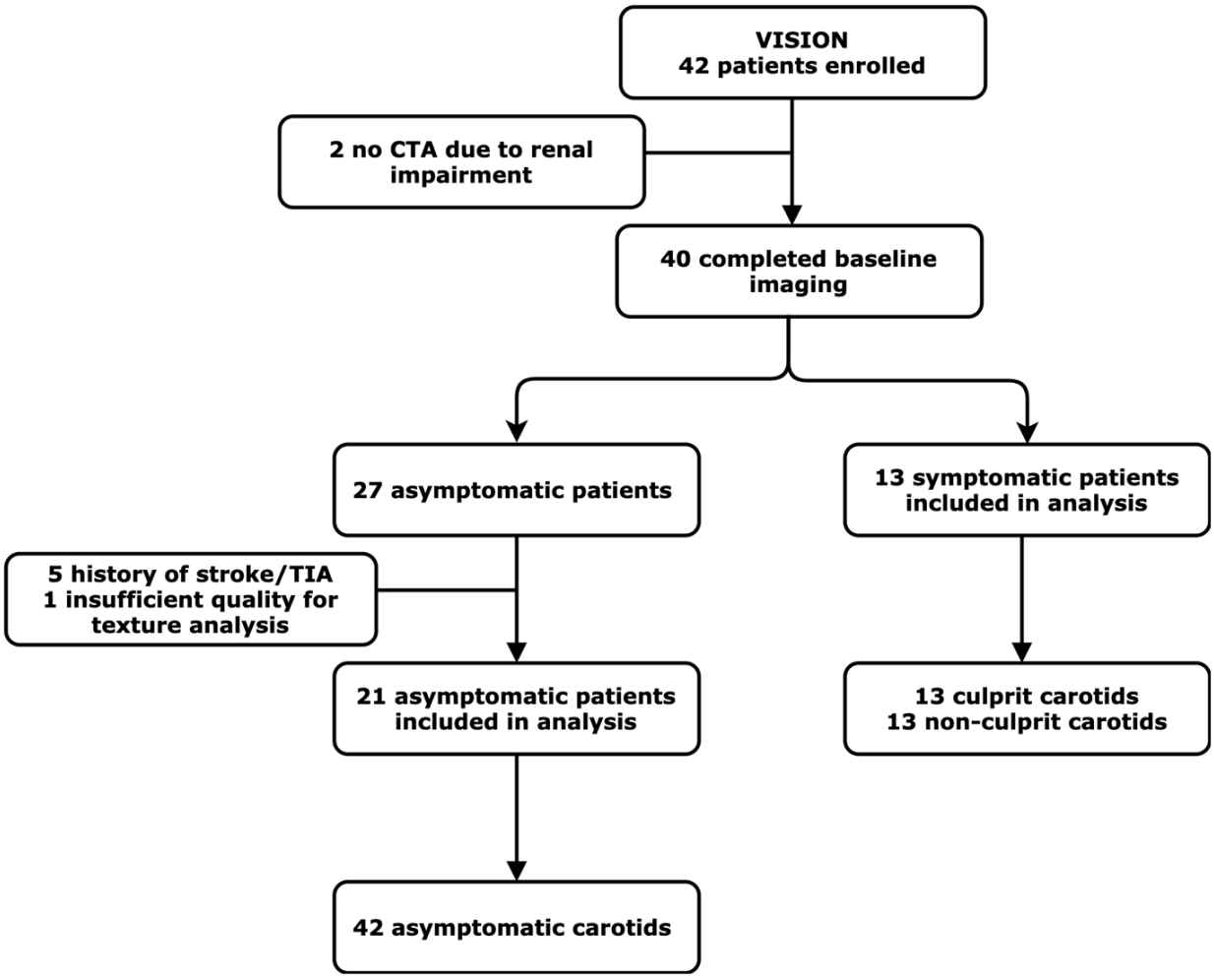


Figure 2.18 VISION: study ran from 2014 to 2015

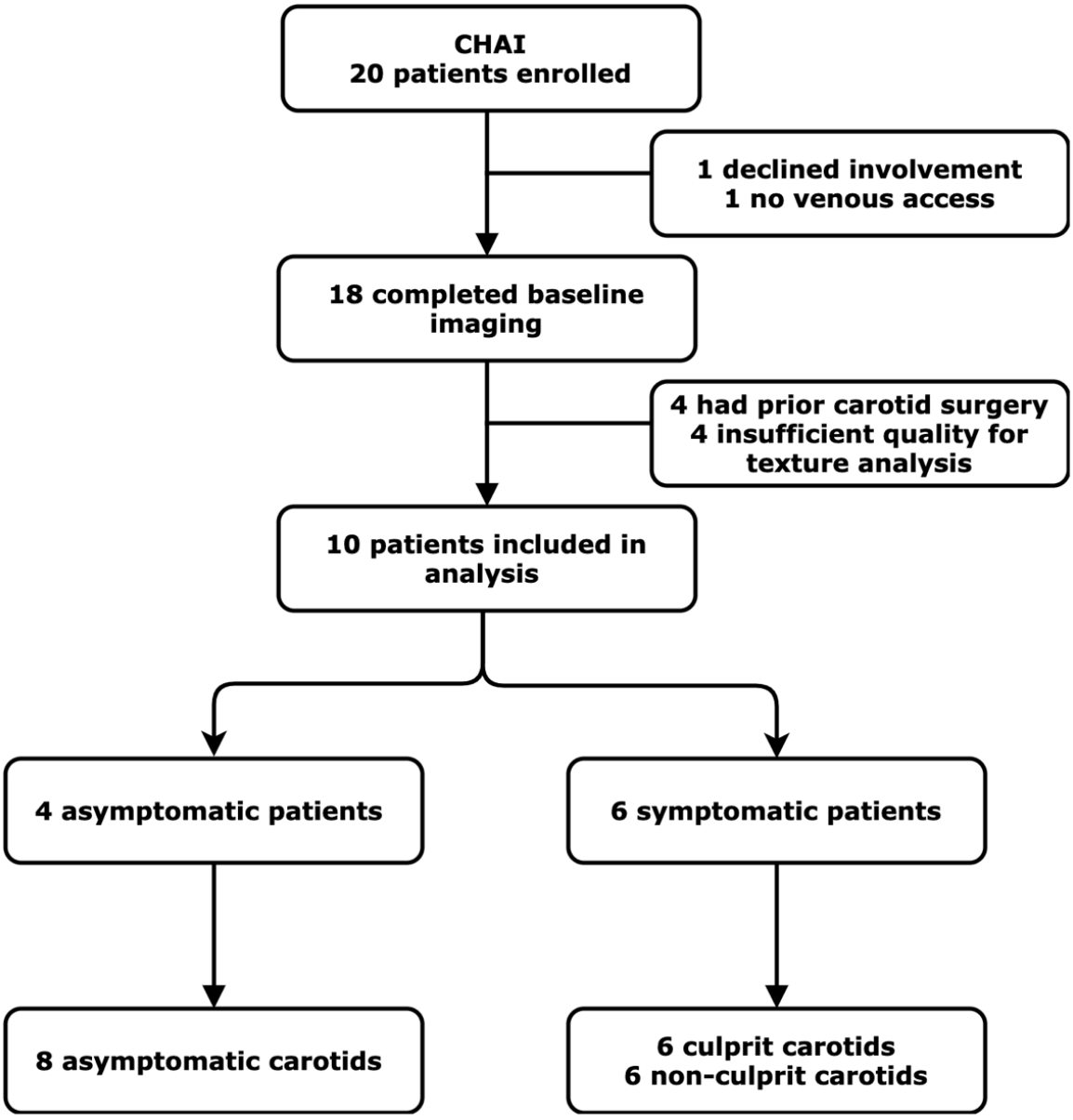


Figure 2.19 CHAI: study ran from 2011 to 2013

2.2.1.1 Image Acquisition

All imaging was performed on a combined GE Discovery 690 PET-CT scanner with an integrated 64-slice CT scanner (GE Healthcare, Waukesha, WI, USA). CTA images were acquired from the aortic arch to the Circle of Willis, using bolus tracking (triggered at 100 HU above baseline) with a region-of-interest placed in the aortic arch and 70-100 ml NIOPAM 300 (Bracco UK Limited, High Wycombe, UK) radiocontrast injected at 5 ml/second, followed by a 50 ml chaser of normal saline. Acquisition parameters: tube voltage 120 kV, maximum tube current 200 mA, rotation time of 0.8 second, pitch 0.969:1, in-plane pixel spacing of minimum 0.30×0.30 mm to maximum 0.59×0.59 mm, slice thickness 0.625 mm with reconstruction interval (i.e., spacing between slices) of 0.4 mm.

Unenhanced CT images were acquired from the arch of the aorta to the external auditory meatus. CT acquisition parameters: tube voltage 120 kV, tube current 40 mA, rotation time 0.5 seconds, pitch 1.375, direct field of view 50 cm, section thickness 3.75 mm reconstructed to 3.27 mm.

2.2.2 Image Quality Assessment
2.2.2.1 Quantitative assessment of image quality

To assess image quality, the mean CT attenuation value in Hounsfield units and the image noise (computed as the standard deviation of the CT attenuation value in HU), were calculated for the main structures in a CTA scan: fat, air, muscle and bone. This was done for 10% of all scans for quality assurance. Regions-of-interest (ROIs) were placed on the different regions: 4 ROIs per region for 5 consecutive slices, with values exported as a .csv file for analysis. This method was recommended and supervised by Dr Fulvio Zaccagna (FZ), Research Associate and Neuroradiologist.

The signal-to-noise ratio (SNR) for the different structures was calculated according to Equation 2.1, in which the mean signal level was taken as the mean attenuation value within the ROI, and the image noise as the standard deviation. SNR is used as a measure of image quality, which compares the level of desired signal to the level of image noise. The higher the SNR, the less noise is present in the image.

$$SNR = \left| \frac{\mu_{sig}}{\sigma_{bg}} \right|$$

Equation 2.1 Signal-to-Noise Ratio (SNR):

μ_{sig} (mean signal level) corresponds to the mean HU of the ROIs and σ_{bg} (image noise) corresponds to the standard deviation of pixel values in the ROIs. | | indicates use of absolute values.

The contrast-to-noise ratio (CNR) was calculated for the carotid arteries of 10% of all CTA scans as a surrogate marker of the quality of the scan in terms of contrast timing. ROIs were placed bilaterally in the carotid arteries and the musculature closest to the carotid arteries[154]. ROIs were marked as large as possible in the vessel lumen whilst being careful to avoid calcifications and severe artefacts caused by dental prostheses or motion. The mean CT attenuation and noise was calculated for individual subjects by averaging the values derived from both sides of the arteries. The CNR was calculated according to Equation 2.2, in which the mean carotid luminal CT attenuation (signal A) minus the CT value of the background

perivascular musculature (signal B), was all divided by the standard deviation of the carotid artery. This method was adapted from Almutairi et al. 2015[154], and Guziński et al.[155]. The SNR was also calculated for the carotid artery, using the mean attenuation HU of the carotid artery divided by the standard deviation in HU of the carotid artery as in Equation 2.1.

$$CNR = \frac{\left|\mu_{sigA} - \mu_{sigB}\right|}{\sigma_{bg}}$$

Equation 2.2 Contrast-to-Noise Ratio (CNR):

μ_{sigA} (mean signal level) corresponds to the mean HU of the ROIs in the carotid arteries; μ_{sigB} (mean signal level) corresponds to the mean HU of the ROIs in the perivascular musculature; σ_{bg} (image noise) corresponds to the standard deviation of pixel values in the ROIs in the carotid arteries. | | indicates use of absolute values.

2.2.2.2 Qualitative assessment of image quality

All CTA and CT scans were assigned a qualitative image quality (IQ) score as an indication of the interpretability of the medical image for the reader. There were four possible grades in the IQ score, described in Table 2.8. This method was adapted from Engel et al. 2012[156]. Any image with a quality score of 0 was prospectively excluded from analysis as the image quality would be insufficient for texture analysis.

Table 2.8 Image Quality (IQ) Score

Image Quality Score	Definitions
3	Excellent: no artefacts. If calcification is present, this does not impinge interpretability.
2	Good: only minor artefacts and/or minor calcification interference.
1	Moderate: artefacts or calcification are noticeable, but image is still interpretable.
0	Poor: uninterpretable due to artefacts and/or heavy calcification

2.2.3 Carotid Characteristics

2.2.3.1 Carotid Calcium Scoring and Calcium Spatial Distributions

Carotid calcification quantification was assessed by using the 'Calcium Scoring' plug-in of OsiriX (Version 10.0.3, Pixmeo SARL, Bernex, Geneva, Switzerland) on unenhanced CT images, as per previous methodology[151]. The calcification detection threshold was based on a CT attenuation value of 130 HU. The plug-in quantifies vascular calcification according to the Agatston method[54], which is commonly used for coronary artery calcification quantification. The Agatston method calculates the calcium score by multiplying the area of calcium by a multiplication factor (see Table 2.9) determined by the highest attenuation value present.

Table 2.9 Agatston Multiplication Factor according to Hounsfield units

Hounsfield units (HU)	Multiplication factor (Agatston scoring)
130-199	1
200-299	2
300-399	3
>400	4

The calcium burden in each carotid artery per patient was analysed for 14 consecutive axial slices, corresponding to the area of interest (approximately 3 cm) for carotid artery texture analysis. The carotid calcium burden is reported on a slice-by-slice basis for calcium distribution analysis and on a total calcium score per artery basis. The total calcium score per artery represents the sum of all single calcium scores from the 14 slices of carotid artery, which includes the common and internal carotid arteries and is expressed as a score in Agatston units (AU). Inter- and intra-observer agreement was assessed by repeated calcium scoring of 10% of the scans by a second reader (NE), and by the same reader (EPVL) respectively, several weeks after the initial analysis with calculation of the appropriate ICC values.

2.2.3.2 Carotid Plaque Type

Carotid lesions were classified according to established coronary CT criteria for plaque composition: calcified, non-calcified, or mixed plaque (contains both calcified and non-calcified plaque)[157]. JT and RS (cardiology registrars) checked 20% of the studies in terms of carotid plaque classification performed by EPVL; any discrepancies were agreed by consensus.

2.2.3.3 Carotid Stenosis Measurements: NASCET

After creating curved multiplanar reformatted images of the carotid artery of interest, the degree of stenosis was quantified using the North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria[158] on CTA images. Training for carotid stenosis measurements was provided by JW (radiologist). The diameter of the stenosed segment was compared with the most distal normal segment of the internal carotid artery and was calculated as a percentage using Equation 2.3:

$$\frac{B - A}{A} \; x \; 100\%$$

Equation 2.3 Measurement of carotid stenosis (NASCET)

where A is the minimum luminal diameter at the point of maximal stenosis and B is the diameter of the normal distal internal carotid artery.

2.2.4 Radiomic Analysis with TexRAD

2.2.4.1 TexRAD Software and Algorithms

TexRAD (Feedback Medical Ltd, Cambridge, UK) is a commercial research texture analysis software that calculates first order statistical features from the image histogram of ROIs drawn on medical images. A detailed description of the software is provided in Chapter 1, section 1.3.2.1.

In TexRAD, 6 histogram-derived features are calculated, these are termed: (1) mean, (2) mean of positive pixels (MPP), (3) kurtosis, (4) skewness, (5) standard deviation and (6) entropy and are further explained in Table 2.10. Please note that within the results section, in order to distinguish between TexRAD texture features and statistical summary measures, TexRAD texture features will be enclosed within ‘’. For example, whereas the mean refers to the average of data points, ‘mean’ refers to the TexRAD texture feature described in Table 2.10.

Table 2.10 TexRAD first order texture feature definitions

Texture Feature	Definitions
Mean	The average CT attenuation of the ROI image histogram
Mean of Positive Pixels	Mean CT attenuation of the ROI image histogram when accounting only Hounsfield units above 0
Kurtosis	A measure of the peakedness of the histogram
Skewness	The asymmetry of the image histogram
Standard Deviation	A measure of how much variation or dispersion exists from the mean value
Entropy	Reflects the randomness of the histogram distribution

ROI, region-of-interest. Adapted from Miles et al. 2013[76]

TexRAD can extract texture features from unfiltered images, that is, the values are derived from the image histogram without any prior filtering applied to the image. TexRAD can also extract texture features from filtered images using an initial Laplacian of Gaussian (LoG) band-pass filter to reduce image noise and enhance edges before the extraction of features from the image histogram.

Different spatial scale filters (SSF) ranging from object radii of 0 (equivalent to unfiltered image), 2, 3, 4, 5, and 6 mm can be used to dictate the size of the LoG filter applied to the image prior to feature extraction in order to extract fine, medium and coarse texture features from the ROI. However, the original parameters for these SSFs were derived for oncological purposes rather than cardiovascular vessels, for example, in oncology, SSF =2 is used to extract fine textures from tumours, SSF=4 for medium-sized textures and SSF=6 for coarse texture features.

If we consider the size of a tumour, for example gastrointestinal stromal tumours can be in the range of centimetres in diameter[159], whereas carotid arteries in axial cross-section are in the range of millimetres[160]. This represents a 10-fold difference in size, therefore, the features derived from different default TexRAD SSFs may not relate to fine, medium and coarse features of carotid artery lesions, see Figure 2.20.

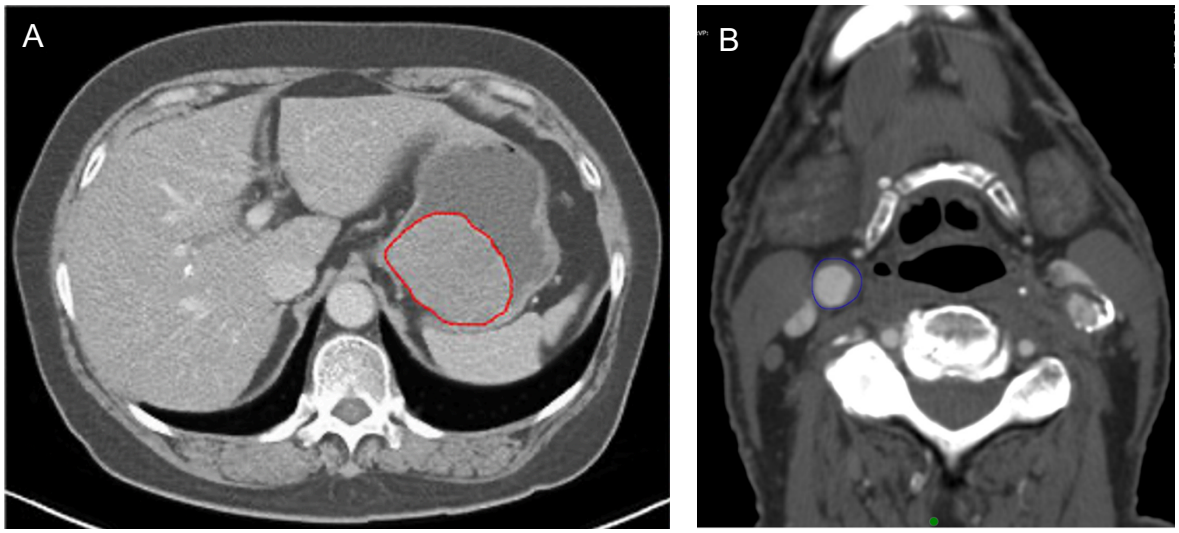


Figure 2.20 Contrast-enhanced CT: Size of tumour versus carotid artery

Figure 2.4A is reproduced from Wang et al. 2019[159] whereby the gastrointestinal stromal tumour was manually delineated at the largest cross-sectional area and the gastric lesion is 6.2 cm in size. Figure 2.4B is a screenshot of the common carotid artery that was manually delineated in TexRAD by EPVL. Krejza et al. measured the mean diameters of the internal carotid artery and common carotid artery (CCA) of 500 patients and found a mean CCA diameter of 6.10 mm in women and 6.25 mm in men[160].

We therefore investigated TexRAD-derived texture features using: (1) no filter and (2) standard 2, 3, 4, 5, and 6 mm SSF filters used in previous TexRAD oncology applications and in FZ's Radiological Society of North America abstract regarding his preliminary study of carotid artery CT texture analysis involving 12 patients with a history or stroke/TIA[161].

The relevant unenhanced CT and CTA images were uploaded to the TexRAD server for analysis. TexRAD texture analysis was performed by a single operator (EPVL) blinded to clinical data, and to the patient and/or culprit carotid artery status. As there was no accepted standard for CT carotid texture analysis at the time, we initially used four different algorithms developed in collaboration with FZ and Dr Balaji Ganeshan (BG, Senior Imaging Scientist and former TexRAD developer) to analyse the images in TexRAD, described in Table 2.11. These algorithms specify the range of Hounsfield units within the ROI that are used in TexRAD texture feature calculations. This process of restricting the range of HU with the restricted algorithms is also known as resegmentation[162].

Table 2.11 TexRAD Algorithms

Algorithm	Range of values (HU)	Purpose
Unrestricted	-1000 to +1000	No calcium or contrast exclusion
Restricted I	0 to 200	Excludes calcium aggressively, limits effect of luminal contrast, perivascular fat exclusion
Restricted II	-1000 to +200	Excludes bone and limits effect of luminal contrast
Restricted III	-1000 to +300	Moderate calcium exclusion

HU, Hounsfield units.

Please note: When using the restricted I algorithm for analysis, which restricts ROI pixels to those between 0 to 200 HU, the TexRAD texture feature 'mean', and the TexRAD texture feature 'mean of positive pixels' have the same value and therefore 'MPP' becomes redundant. Therefore, for further statistical analysis of TexRAD texture features using the restricted I algorithm, 'MPP' was removed due to duplication and redundancy of information.

2.2.4.2 Single-slice analysis and Multi-slice analysis rationale

In oncological texture analysis and radiomic studies, single-slice and multi-slice approaches have been investigated. For example, Cecco et al. used a single-slice approach to analyse pre-treatment and mid-treatment MRI images from rectal cancer patients[163]. The axial slice with the largest cross-sectional area of the tumour was used for analysis whereby the tumour on that slice was manually segmented to create an ROI from which TexRAD features were extracted. The single-slice approach was sufficient to detect differences in TexRAD texture features between the pre-treatment and mid-treatment MRI scans. Ng et al. investigated the use of the single-slice approach versus the multi-slice approach (i.e. the largest cross-sectional area versus whole tumour analysis) on texture analysis and clinical outcome prediction. They found that whole tumour analysis was more representative of tumour heterogeneity[164].

Zaccagna et al. in his preliminary carotid CTA texture analysis study extracted texture features from ROIs drawn on a single axial slice per carotid artery, selected to best represent the carotid bifurcation[161]. Furthermore, in the original cardiovascular imaging research projects (ICARUSS[151], VISION[152] and CHAI[153]) from which the carotid CT imaging dataset was derived (see section 2.2.1), the patients had corresponding PET imaging of the carotid arteries to investigate the diagnostic utility of different PET tracers such as ^{18}F-fluorodeoxyglucose (FDG), ^{18}F-sodium fluoride (NaF), ^{18}F-fluoromisonidazole (FMISO) and/or ^{68}Ga-DOTA-[Tyr3]octreotate (DOTATATE). In these studies, a multi-slice approach was taken to involve 14 consecutive slices of the carotid artery.

The advantage of a single-slice approach is that it is faster and easier to implement within the clinical workflow; for each carotid artery only one ROI needs to be manually delineated. However, a single-slice approach may not capture sufficient information to represent the carotid artery lesion and a multi-slice approach may be better for predicting clinical outcomes. Multi-slice analysis captures the texture data from multiple slices across the volume. We therefore investigated both approaches to see whether a single-slice approach was sufficient to capture discriminative information between different carotid artery types and whether predictive ability could be improved using a multi-slice approach.

2.2.4.3 TexRAD Single-Slice Analysis: Carotid Bifurcations

The heaviest burden of carotid atherosclerosis tends to be within 2 cm of the carotid artery bifurcation. In single-slice analysis, bilateral carotid arteries were evaluated in each patient from a single axial CTA slice of original slice thickness 0.625 mm and slice spacing of 0.4 mm, located at the carotid bifurcation as a standardized landmark. ROIs were drawn within the TexRAD software to encompass the whole vessel, including the outer wall. ROIs drawn using TexRAD were downloaded and saved in Extensible Markup Language (XML) format, so that they could be used with PyRadiomics for direct comparison of texture parameters in later analysis (see chapter 3), and for analysis of the corresponding unenhanced CT carotid artery axial slice. The CTA images were analysed with the reviewer (EPVL) blinded to the subjects' demographic information and clinical status.

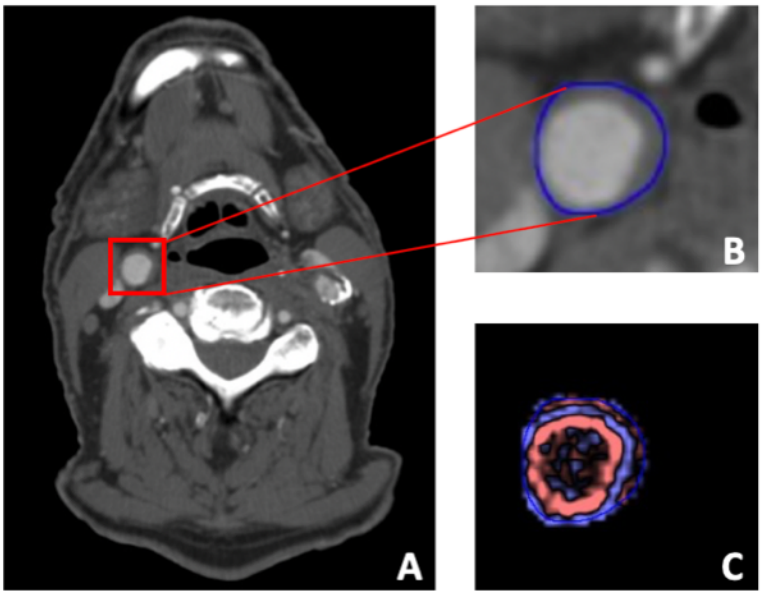


Figure 2.21 TexRAD Feature Extraction

A. Carotid computed tomography angiography
B. Manual ROI segmentation performed in TexRAD
C. Colour map of filtration-histogram texture feature extraction in TexRAD

2.2.4.4 TexRAD Multi-Slice Analysis

For multi-slice analysis, CT angiography and unenhanced CT images were resampled to 3 mm slice thickness using the OsiriX MD software resampling plugin (Version 10.0.3, Pixmeo SARL, Bernex, Geneva, Switzerland) as per published methods[151–153]. A 3 mm slice thickness was chosen to remain consistent with previous PET-CT imaging studies conducted in ICARUSS[151], VISION[152] and CHAI[153].

An ROI was manually drawn on a single axial slice around the whole vessel and automatically propagated in TexRAD for multi-slice analysis along the common carotid artery and internal carotid artery. Manual adjustments of the ROIs were made throughout the volume where necessary to ensure the ROI included the outer wall of the carotid artery. 14 carotid artery slices were captured in the analysis bilaterally, with the carotid bifurcation designated as slice 0, as described in previous vascular imaging studies[151–153]. The slices started from slice -3 to slice +10, i.e. three slices below the carotid bifurcation to 10 slices above, as shown in Figure 2.22.

TexRAD has a function specifically for multi-slice analysis within its data miner, which takes into account all of the voxel values in the multi-slice carotid artery ROIs to generate a cumulative histogram from which first order texture features can be derived. This TexRAD function was used to calculate the first order 3D features of the carotid arteries from the 14 consecutive slices that were delineated. The features were: (1) ‘mean’, (2) ‘mean of positive pixels’ (MPP), (3) ‘standard deviation’, (4) ‘entropy’, (5) ‘skewness’ and (6) ‘kurtosis’.

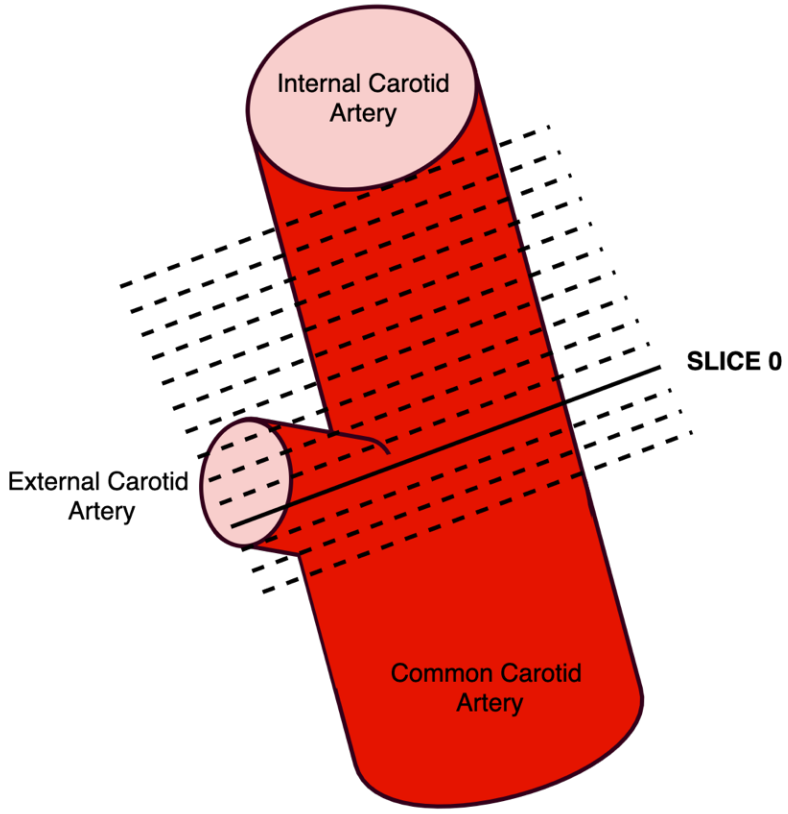


Figure 2.22 Depiction of Carotid Multi-slice Analysis

Slice 0 corresponds to the carotid bifurcation. Multi-slice analysis involved 14 consecutive slices from slice -3 to slice +10, capturing parts of the common and internal carotid artery.

2.2.4.5 TexRAD Subset Analysis

2.2.4.5.1 Comparisons with asymptomatic carotid plaque only

Since the asymptomatic patients in this study came from three possible categories (described in section 2.2.1), not all of the 50 asymptomatic carotid arteries had evidence of carotid atheroma. 22 asymptomatic carotid arteries had no evidence of carotid plaque leaving 28 carotid arteries with proven asymptomatic atherosclerosis. To explore whether the presence or absence of visible carotid atheroma in asymptomatic carotid arteries impacted the TexRAD texture feature differences with culprit or non-culprit carotid arteries respectively, two levels of analyses were conducted: (1) inter-patient comparisons using all asymptomatic carotid arteries (including those without carotid plaque, a total of 50 asymptomatic carotid arteries) versus 41 culprit and 41 non-culprit carotid arteries respectively and (2) inter-patient comparisons using only asymptomatic carotid arteries with carotid plaque (28 carotid arteries) versus 41 culprit and 41 non-culprit carotid arteries respectively. These analyses were conducted for both single-slice and multi-slice approaches.

2.2.4.5.2 Carotid Plaque Type

Amongst the carotid arteries (asymptomatic, culprit and non-culprit), different carotid plaque types were present: calcified, noncalcified and mixed. A subset analysis in the multi-slice approach was therefore conducted to explore TexRAD texture features stratified by carotid plaque type to control for these differences.

2.2.4.5.3 Carotid Stenosis Severity

According to the NASCET criteria, carotid stenosis can be classified as mild (below 50%), moderate (50-69%) or severe (70-99%)[165]. The National Institute for Health and Care Excellence (NICE) guidelines recommend that patients with non-disabling stroke or TIA and carotid stenosis of 50-99% (NASCET criteria), should be considered for urgent carotid endarterectomy[166]. Symptomatic patients with less than 50% stenosis should not have surgery but rather receive medical therapy. Due to the differences in patient management between carotid stenosis below 50% and above 50%, subset analysis of TexRAD texture features was conducted for these two categories of carotid stenosis severity.

2.2.5 Unenhanced CT Analysis

To explore the effects of contrast in the carotid artery lumen on the radiomic features, corresponding unenhanced CT images were analysed. The unenhanced CT images were non-rigidly co-registered with their respective CTA images using anatomical landmarks in OsiriX. Previously, the ROIs drawn for the CTA images had been saved as XML files. These could be reuploaded to TexRAD for the unenhanced CT images of the same patient and manually adjusted if necessary, to encompass the carotid artery. The unenhanced CT images were subsequently analysed according to the methodology previously described for TexRAD (section 2.2.4.4) for multi-slice analysis.

2.2.6 Statistical Analysis

Continuous data were checked for normality by plotting histograms for visual inspection and by using the Shapiro-Wilk test. Descriptive statistics are reported as mean (standard deviation [SD]) for continuous variables or median (interquartile range [IQR]) as appropriate.

For statistical comparisons between culprit versus non-culprit carotid arteries (intra-patient comparisons) within symptomatic patients, the paired t-test or the Wilcoxon signed-rank test was used as appropriate after assessing for normality of differences. To compare between asymptomatic and symptomatic patients, the unpaired t-test or the non-parametric Mann-Whitney *U* test was used as appropriate. To compare proportions within symptomatic patients, McNemar's test on paired proportions was used. For inter-patient comparisons, Pearson's Chi-squared test or Fisher's Exact test was used where appropriate.

For radiomic feature comparisons however, and radiomic feature reporting – since there are a large number of parameters and there was a mixture of normally and non-normally distributed variables and differences between the groups, the non-parametric tests were used in each case and the radiomic features were reported using median (IQR). To compare between asymptomatic and symptomatic patients, the nonparametric Mann-Whitney *U* test was used for the following inter-subject comparisons of radiomic features: (1) asymptomatic carotid versus culprit carotid (Asx vs CC) and (2) asymptomatic carotid versus non-culprit carotid (Asx vs NC). P-values <0.05 were considered significant, with correction for multiple comparisons where appropriate (Benjamini-Hochberg procedure).

To measure the intra-observer agreement in calcium scoring, the one-way random-effects model intraclass correlation coefficient (ICC) for absolute agreement was calculated. To assess inter-observer agreement in calcium scoring, the two-way random-effects model ICC for absolute agreement[167] was calculated. To compare the level of agreement between texture features derived from CT angiography scans and unenhanced CT, the intraclass correlation coefficient for absolute agreement and consistency were calculated. An ICC $<$ 0.5 was considered poor agreement[168], between 0.5 to 0.9 was moderate agreement and ICC $\geq$ 0.9 was considered excellent agreement. ICC values for single measures are reported.

All statistical analyses were performed in IBM SPSS Statistics for Macintosh (Version 25.0. Armonk, NY, USA: IBM Corp.), Microsoft Excel for Mac (Version 16.37), R version 4.0.2 and Python version 3.7.4.

2.3 Results

2.3.1 Quantitative and Qualitative Image Quality Assessment

2.3.1.1 Quantitative Image Quality Assessment

10% of CTA scans (n=7) were randomly selected and evaluated for quantitative image quality, ensuring suitable representation of scans from all three studies (ICARUSS, CHAI and VISION). Figure 2.23 illustrates the process of ROI placement. The SNR for different structures is reported in
Table 2.12.

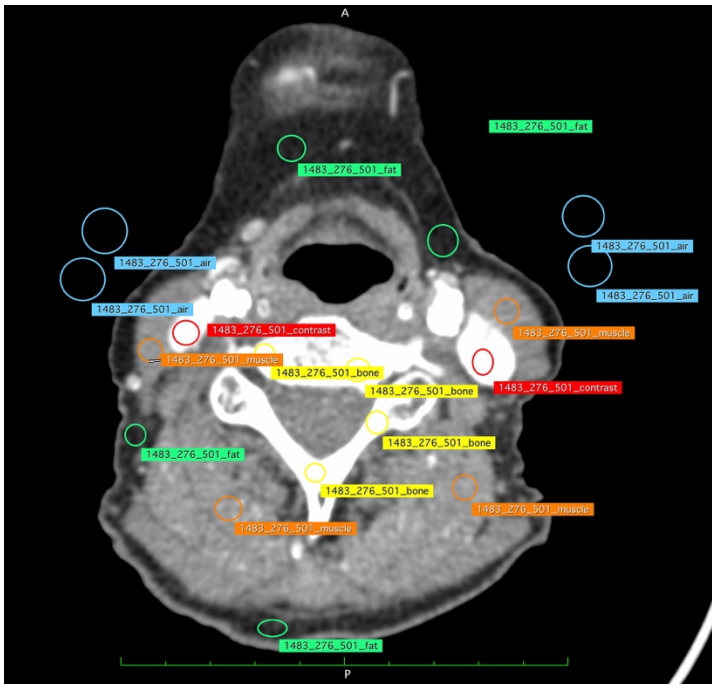


Figure 2.23 ROI placement for quantitative image quality analysis.

Axial CT angiography image of the neck with 4 ROIs placed on different structures to represent air, fat, bone and muscle. This was completed for 5 consecutive slices per study. ROIs are the coloured circles.

Table 2.12 Quantitative Image Quality Assessment of Fat, Air, Muscle and Bone

Structures	CT attenuation level (HU)	Image Noise (HU)	SNR
Fat	-98.81 ± 16.10	19.08 ± 5.24	5.72 ± 2.40
Air	-991.90 ± 5.24	8.22 ± 2.11	128.24 ± 35.11
Muscle	67.67 ± 10.92	13.39 ± 2.34	5.09 ± 0.65
Bone	672.74 ± 144.95	216.37 ± 36.67	3.17 ± 0.83

Values are reported as mean ± S.D; HU, Hounsfield units; SNR, Signal-to-noise ratio

14 carotid arteries were assessed from 7 random CTA scans, as well as the surrounding perivascular muscle to calculate SNR and CNR values as an indication of contrast timing and imaging quality. These results are reported in Table 2.13. The SNR and CNR values for the carotid arteries are comparable with those in reported CTA studies of peripheral vascular beds[154], and provides quality assurance for the use of these CTA scans for carotid artery radiomic analysis.

Table 2.13 Quantitative Image Quality Assessment of Carotid Arteries

Structures	CT attenuation level (HU)	Image Noise (HU)	SNR	CNR
Carotid Artery	351.10 ± 74.75	24.65 ± 6.43	15.60 ± 6.36	12.71 ± 5.67
Perivascular Muscle	67.45 ± 10.47	14.44 ± 3.71	4.87 ± 1.04	-

Values are reported as mean ± S.D; HU, Hounsfield units; SNR, Signal-to-noise ratio; CNR, Contrast-to-noise ratio.

2.3.1.2 Qualitative Image Quality Assessment

All eligible CTA scans (74 scans) were qualitatively assigned an image quality score by a single reader, EPVL, according to the criteria in Table 2.8. Scans with an image quality score of 0 were deemed insufficient quality for radiomic analysis due to excessive calcification or other imaging artefacts. Overall, 8 CTA scans received an image quality score of 0 and were excluded from the final analysis. There were 66 remaining CTA scans that were suitable for texture analysis. The distribution of image quality scores is shown in Figure 2.24.

Figure 2.1 Distribution of CTA Image Quality Scores

CTA, computed tomography angiography

2.3.2 Study population

A total of 66 patients were included in the final analysis following image quality assessment. This consisted of 41 symptomatic patients (82 carotid arteries: 41 culprit carotids; 41 non-culprit carotids) and 25 asymptomatic patients (50 asymptomatic carotids). The mean age of the study population was 71.4 (SD 9.2 years), with 75.8% males. Of the 41 symptomatic patients, 30 (73.2%) had a stroke and 11 (26.8%) had a TIA. 65.9% (27/41) of culprit carotids occurred on the right-hand side and 34.1% of culprit carotids (14/41) occurred on the left-hand side. Table 2.14 provides a summary of the clinical characteristics for all patients in this study.

Table 2.14 Patient Characteristics: Symptomatic versus Asymptomatic Patients

Characteristics	All Patients	Asymptomatic	Symptomatic	p-value
n	66	25	41	-
Age (years), Mean ± SD	71.4 ± 9.2	66.9 ± 8.7	74.1 ± 8.4	0.001*
Male - *n* (%)	50 (75.8)	18 (72.0)	32 (78.0)	0.578
Hypertension - *n* (%)	39 (59.1)	12 (48.0)	27 (65.9)	0.152
Diabetes Mellitus - *n* (%)	11 (16.7)	3 (12.0)	8 (19.5)	0.427
Smoking History (current/former) – *n* (%)	45 (68.2)	16 (64.0)	29 (70.7)	0.569
Current statin - *n* (%)	49 (74.2)	24 (96.0)	25 (61.0)	0.001*

*Unpaired t-test used to compare continuous parametric data; Pearson's Chi-squared test used to compare proportions; two-sided p-values <0.05 was deemed statistically significant; n, number of patients; SD, standard deviation; *p-value<0.05*

The asymptomatic patients were younger than symptomatic patients (p=0.001), and they were more likely to be taking a statin at the time of imaging than symptomatic patients (p=0.001). Otherwise, there were no statistically significant differences in other cardiovascular risk factors (male gender, hypertension, diabetes mellitus or smoking history) between asymptomatic and symptomatic patients.

2.3.3 Carotid Characteristics
2.3.3.1 Carotid Calcium Score

Inter-observer reliability of carotid calcium scoring was assessed for 34 carotid arteries with NE's (stroke registrar) ratings and EL's ratings. The ICC value (two-way random effects model for absolute agreement) was 0.952 (0.906, 0.976) for total calcium score, indicating excellent agreement. Intra-observer reliability was assessed for 34 carotid arteries, the ICC value (one-way random effects model for absolute agreement) was 0.996 (0.993, 0.998).

Table 2.15 Calcium Scores in Asymptomatic, Culprit and Non-culprit Carotid Arteries

Carotid Calcium Score (AU)	**Asymptomatic Carotids (n=50)**	**Culprit (n=41)**	**Non-culprit (n=41)**	**Comparisons**	***p-value***
Median (IQR)	3 (0, 151)	263 (95, 701)	387 (63, 659)	Asx vs CC	<0.0001*
				Asx vs NC	<0.0001*
				CC vs NC	0.706
Minimum	0	0	0	-	-
Maximum	1898	1963	1671	-	-
Zero CAC- n (%)	25 (50.0%)	1 (2.4%)	4 (9.8%)	Asx vs CC	<0.0001*
				Asx vs NC	<0.0001*
				CC vs NC	0.250

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; AU, Agatston units; IQR, interquartile range; n, number of carotid arteries. Calcium score comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test, CC vs NC: Wilcoxon signed-rank test; Zero CAC: Asx vs CC or Asx vs NC: Fishers' Exact test where appropriate; CC vs NC: McNemar's test on paired proportions. *p-value<0.05.*

There were no statistically significant differences in total calcium score (AU) between culprit versus non-culprit carotid arteries in symptomatic individuals (p=0.706). Asymptomatic carotids however, had a statistically significantly lower total calcium score than symptomatic carotids (culprit or non-culprit carotids, p<0.0001), as shown in Figure 2.25. A zero carotid calcium score was more frequent in asymptomatic patients than symptomatic patients (p<0.0001).

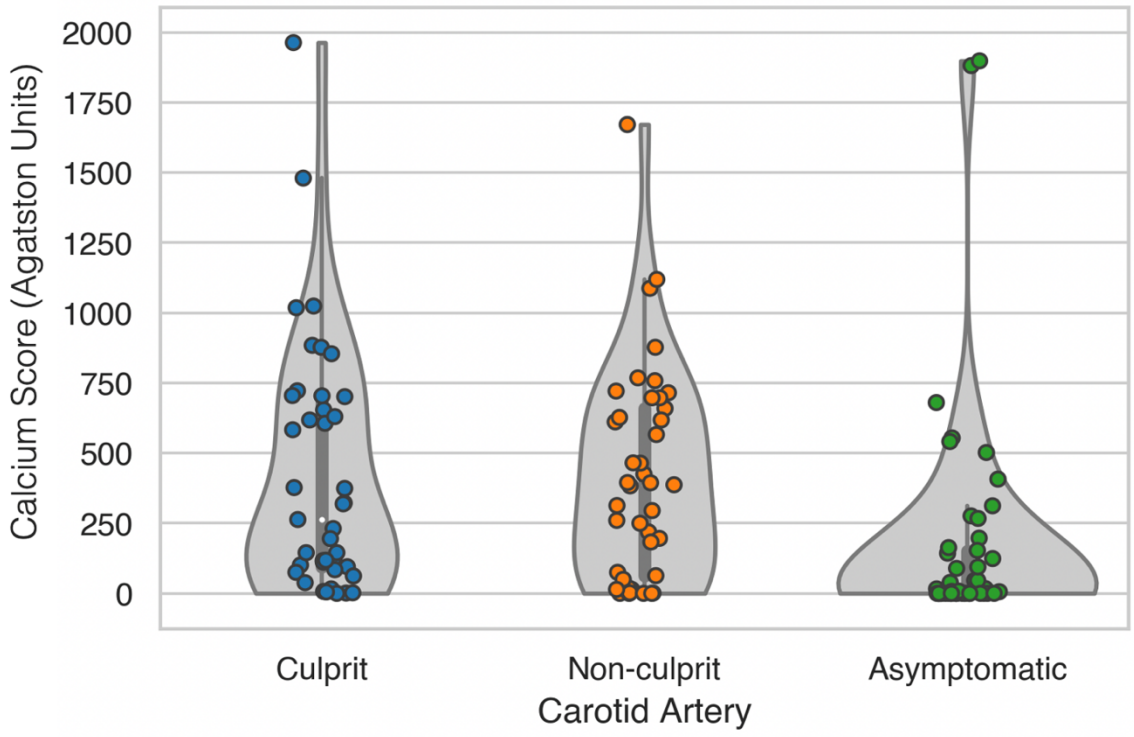


Figure 2.25 Violin plot with strip plot overlay of calcium score distributions in different carotid arteries

Violin plots are a method of displaying data across different categories, providing more information than box-plots alone. The violin plots each contain a dark grey box-plot of the calcium scores corresponding to actual data points and a lighter grey shape that provides a visualisation of the underlying distribution of calcium scores via kernel density estimations. In this figure, the strip plot overlay displays the individual calcium scores and the violin plots demonstrates the distribution of the calcium scores for culprit, non-culprit and asymptomatic carotid arteries respectively.

2.3.3.2 Carotid Calcium Spatial Distribution

The calcium score of individual slices in the 14-slice multi-slice approach of each carotid artery was recorded for symptomatic and asymptomatic patients. With the carotid bifurcation designated as slice 0, we can visualise the spatial distribution of calcium along the length of the analysed carotid artery. As shown in Figure 2.26 and Figure 2.27, the calcium burden was localised around the carotid bifurcation in both culprit and non-culprit arteries (around slice 0). The median values were plotted because calcium score was non-normally distributed. The distribution of calcium in culprit and non-culprit carotid arteries followed the same pattern. There was no statistically significant difference in total calcium score between the culprit and non-culprit carotid arteries in symptomatic patients.

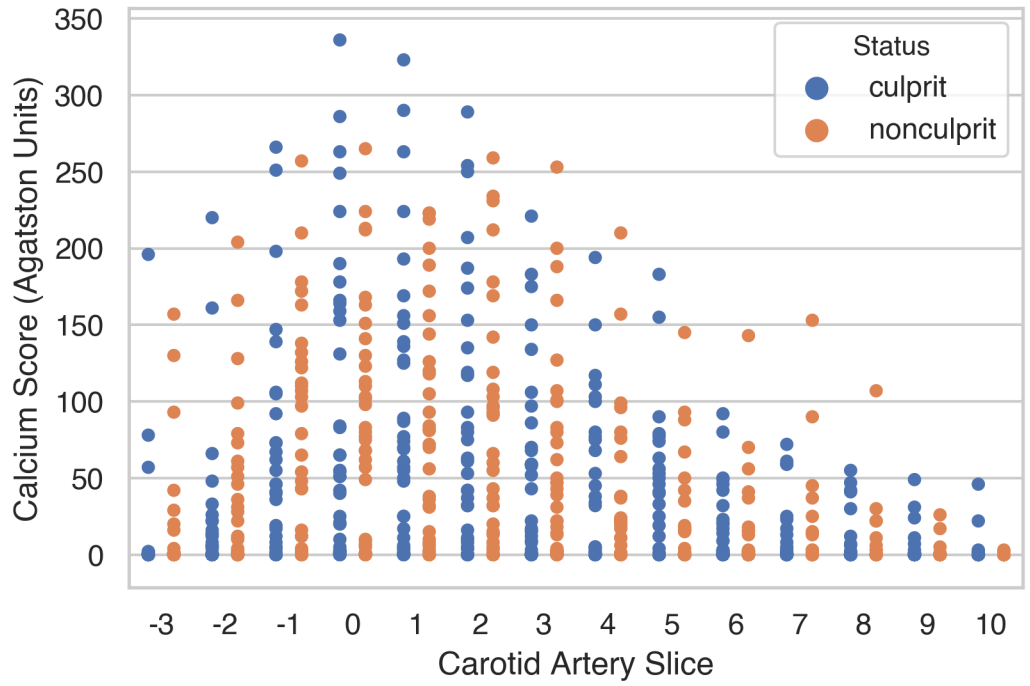


Figure 2.26 Strip plot of individual calcium score values across 14 slices of culprit versus non-culprit carotid arteries

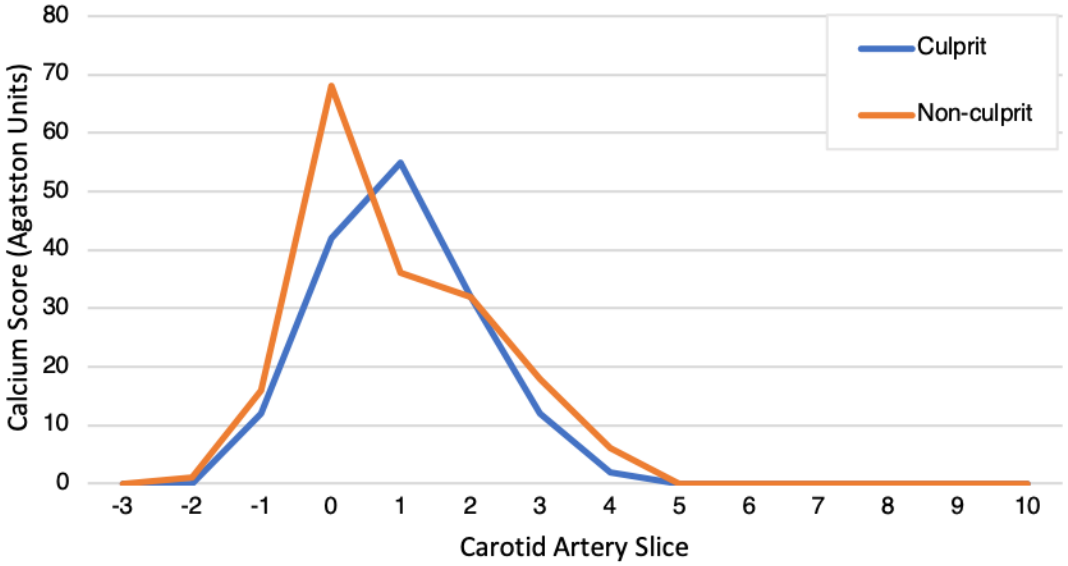


Figure 2.27 Median calcium distribution of culprit versus non-culprit carotid arteries

This figure plots the median calcium score per slice for culprit versus non-culprit carotid arteries, averaging the datapoints displayed in Figure 2.26.

On a slice-by-slice basis, Wilcoxon-signed rank tests revealed no differences in calcium score between culprit and non-culprit carotid arteries (Table 2.16).

Table 2.16 Calcium Score slice-by-slice analysis: Culprit versus Non-culprit

Carotid Slice Number	Calcium Score (AU) – median (IQR)		*p*-value
	Culprit Carotid	Non-culprit Carotid	
Slice -3	0 (0, 1)	0 (0, 1)	0.298
Slice -2	0 (0, 13)	1 (0, 36)	0.062
Slice -1	12 (0, 67)	16 (0, 110)	0.293
Slice 0	42 (2, 153)	68 (1, 113)	0.763
Slice 1	55 (3, 127)	36 (1, 118)	0.544
Slice 2	32 (0, 93)	32 (0, 97)	0.925
Slice 3	12 (0, 59)	18 (0, 73)	0.636
Slice 4	2 (0, 53)	6 (0, 24)	0.18
Slice 5	0 (0, 46)	0 (0, 16)	0.121
Slice 6	0 (0, 17)	0 (0, 2)	0.122
Slice 7	0 (0, 4)	0 (0, 1)	0.327
Slice 8	0 (0, 1)	0 (0, 0)	0.423
Slice 9	0 (0, 0)	0 (0, 0)	0.306
Slice 10	0 (0, 0)	0 (0, 0)	0.287

AU, Agatston units; IQR, interquartile range. The p-value was obtained using the Wilcoxon signed-rank test, after assessing for normality of differences. The slices refer to the 14 consecutive carotid artery slices analysed in a multi-slice approach, with the carotid bifurcation designated as slice 0. The slices started from slice -3 to slice + 10, i.e. three slices below the carotid bifurcation to 10 slices above.

In asymptomatic patients, the median total carotid calcium score was 3 (IQR 151), however, on a slice-by-slice basis, several slices had a median of zero. These low median calcium score values were due to there being 50% of asymptomatic carotid arteries having a zero-calcium score. Figure 2.28 shows a strip plot of the individual calcium score values for each asymptomatic carotid artery across the 14 consecutive CT slices analysed.

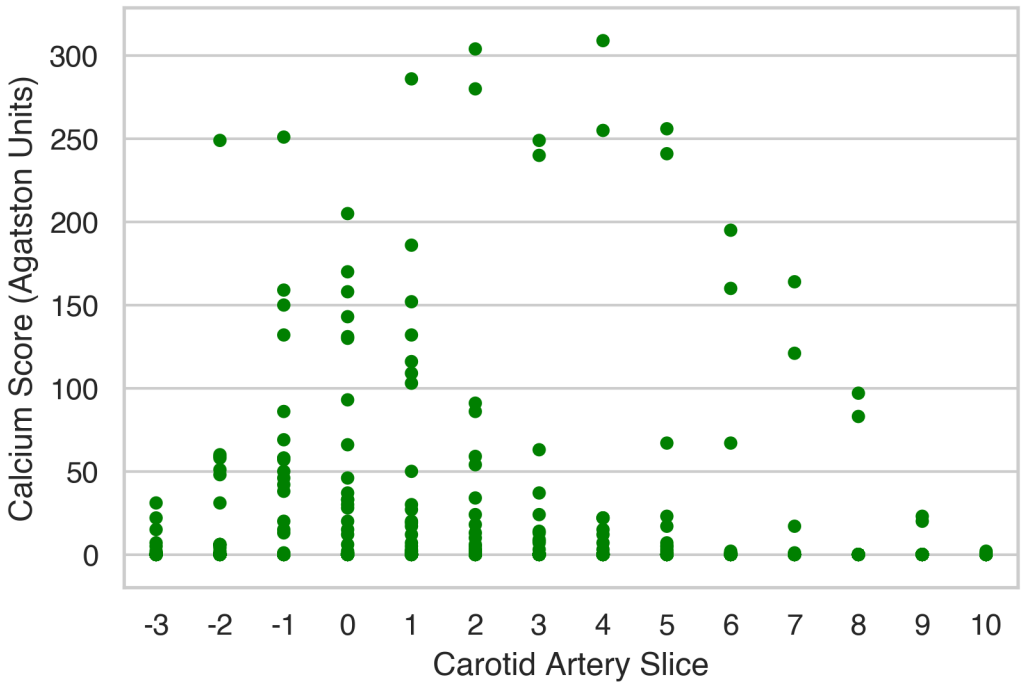


Figure 2.28 Strip plot of individual calcium score values across 14 slices of asymptomatic carotid arteries

Excluding the asymptomatic zero calcium cases allowed for plotting of the median carotid calcium scores for comparison with the culprit and non-culprit carotid arteries. Figure 2.29 demonstrates a similar calcium spatial distribution in asymptomatic patients as in symptomatic patients, whereby calcification builds up around the carotid arteries.

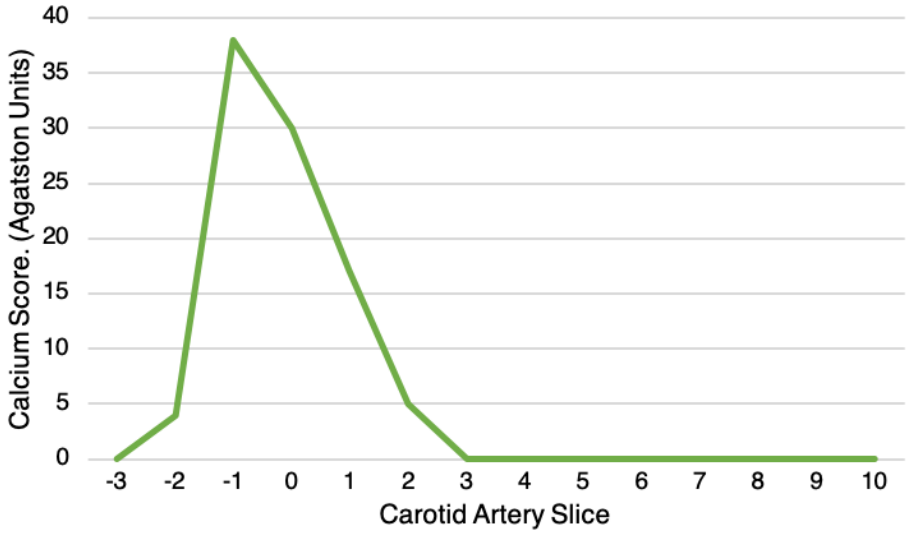


Figure 2.29 Median Calcium Distribution of Asymptomatic Carotid Arteries

A plot of central tendency (mean value) and 95% confidence intervals indicated that in the presence of carotid calcification, the pattern of calcium distribution in asymptomatic and symptomatic patients was similar. There was a tendency for calcium to cluster around the area of the carotid bifurcation (slice 0), shown in Figure 2.30.

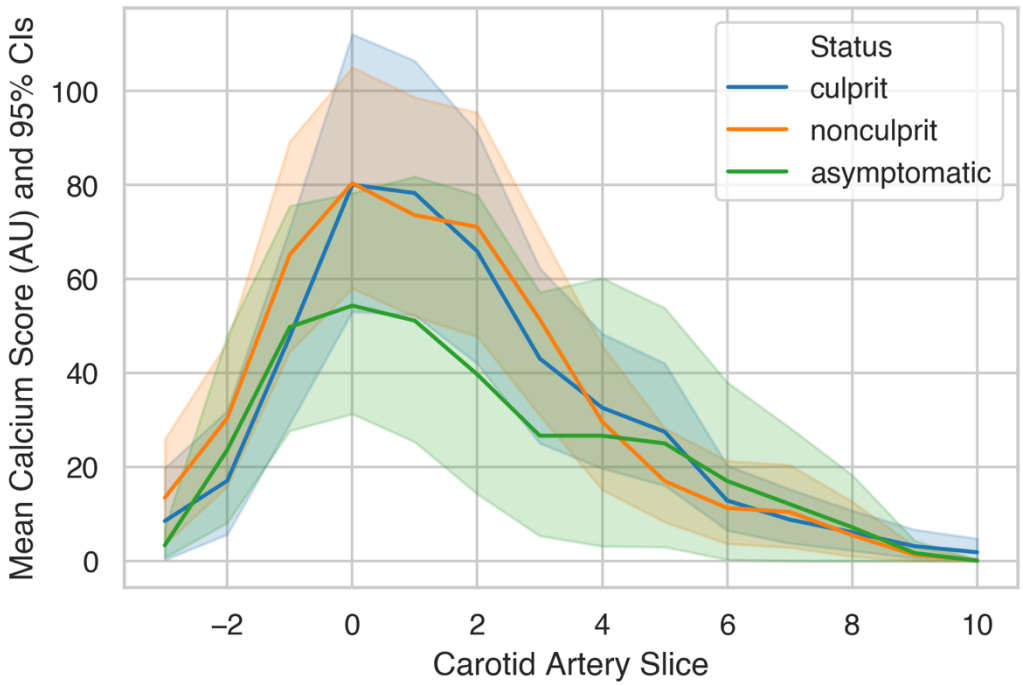


Figure 2.30 Plot of Central Tendency of Carotid Calcium Score with 95% Confidence Intervals

AU, Agatston units; CI, confidence intervals.

2.3.4 Carotid Plaque Characteristics

44% of asymptomatic carotid arteries had no carotid plaque (n=22). In the asymptomatic carotid arteries that had carotid plaque (n=28), 28.0% were calcified plaque types which was similar to the prevalence of calcified plaques in culprit (22.0%) and non-culprit (36.6%) carotid arteries. Culprit carotid arteries were more likely to have the mixed plaque type (73.2%) compared with asymptomatic carotid arteries (26.0%, p<0.0001). Non-culprit carotid arteries were more likely to have mixed plaque as well (51.2%, p=0.013) compared with asymptomatic carotids. There was no difference in the prevalence of noncalcified plaque between asymptomatic, culprit or non-culprit carotid arteries, see Table 2.17.

Table 2.17 Carotid Plaque Characteristics

Carotid Characteristic	Asymptomatic Carotids (n=50)	Culprit (n=41)	Non-culprit (n=41)	Comparisons	*p-value*
Calcified - n (%)	14 (28.0%)	9 (22.0%)	15 (36.6%)	Asx vs CC	0.509
				Asx vs NC	0.382
				CC vs NC	0.070
Noncalcified - n (%)	1 (2.0%)	2 (4.9%)	5 (12.2%)	Asx vs CC	0.587
				Asx vs NC	0.087
				CC vs NC	0.250
Mixed - n (%)	13 (26.0%)	30 (73.2%)	21 (51.2%)	Asx vs CC	<0.0001*
				Asx vs NC	0.013*
				CC vs NC	0.012*
No Plaque	22 (44%)	0 (0.0%)	0 (0.0%)	Asx vs CC; Asx vs NC	<0.0001*

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries. Plaque type comparisons: Asx vs CC or Asx vs NC: Pearson's Chi-squared test or Fishers' Exact test where appropriate; CC vs NC: McNemar's test on paired proportions. N, number of carotids; *p-value<0.05.*

Table 2.18 Carotid Stenosis

Carotid Stenosis	Asymptomatic Carotids (n=50)	Culprit (n=41)	Non-culprit (n=41)	Comparisons	*p-value*
Mean - % stenosis (SD)	10 (20)	72 (17)	40 (22)	Asx vs CC	<0.0001*
				Asx vs NC	<0.0001*
				CC vs NC	<0.0001*
Minimum - % stenosis	0	29	3	-	-
Maximum - % stenosis	97	99	88	-	-

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; n, number of carotids; SD, standard deviation. Carotid stenosis comparison: Asx vs CC or Asx vs NC: unpaired t-test; CC vs NC: paired Student t-test. * p-value<0.05.*

The mean degree of carotid luminal stenosis was higher in culprit carotid arteries versus non-culprit carotid and versus asymptomatic carotid arteries respectively (mean stenosis: CC 72%, NC 40%, Asx 10%, p<0.0001), see Table 2.18. Table 2.19 provides a summary of carotid stenosis categories for the different carotid artery types.

Table 2.19 Carotid Stenosis Categories

Carotid Stenosis	Asymptomatic Carotids (n=50)	Culprit (n=41)	Non-culprit (n=41)	Comparisons	*p-value*
Normal - n (%)	22 (44)	0 (0)	0 (0)	Asx vs CC; Asx vs NC	<0.0001*
1 to <50% - n (%)	25 (50)	3 (7)	27 (66)	Asx vs CC	<0.0001*
				Asx vs NC	0.128
				CC vs NC	<0.0001*
50 to <70% - n (%)	0 (0)	12 (29)	11 (27)	Asx vs CC	<0.0001*
				Asx vs NC	<0.0001*
				CC vs NC	1.000
≥70% - n (%)	3 (6)	26 (63)	3 (7)	Asx vs CC	<0.0001*
				Asx vs NC	1.000
				CC vs NC	<0.0001*

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; n, number of carotids. Frequency comparisons: Asx vs CC or Asx vs NC: Pearson's Chi-squared test or Fishers' Exact test where appropriate; CC vs NC: McNemar's test on paired proportions. * p-value<0.05.*

Culprit carotid arteries were more likely to have severe carotid stenosis (defined as ≥*70%* luminal stenosis) than non-culprit and asymptomatic carotid arteries (p<0.0001). There was no difference between non-culprit and asymptomatic carotid arteries however (7% versus 6% respectively). No asymptomatic carotid arteries had stenosis in the 50 to <70% category, this was more frequent in culprit and non-culprit carotids (29% and 27%). The majority of asymptomatic carotids had a degree of carotid stenosis less than 50% (50%), this was also the case for non-culprit carotid arteries (66%), see Table 2.19.

2.3.5 Carotid Bifurcation Identification

To identify the carotid bifurcation, we first identified the slice at which we could see two distinct vessels: the internal carotid artery (ICA) and the external carotid artery (ECA). For slices of 3 mm slice thickness (as in multi-slice analysis), the carotid bifurcation was taken as the slice below the distinct vessel slice. For slices of 0.625 mm slice thickness (as in single-slice analysis), the carotid bifurcation was taken as the slice three slices below the distinct vessel slice. ICC values for identification of the distinct vessel slice (n=16) between EPVL and NE came to 1.000 (0.999, 1.000) using the two-way random effects model with absolute agreement.

Figure 2.31 demonstrates an example of a slice identified as the carotid bifurcation and the subsequent manual ROI drawn around the adventitia of the carotid artery.

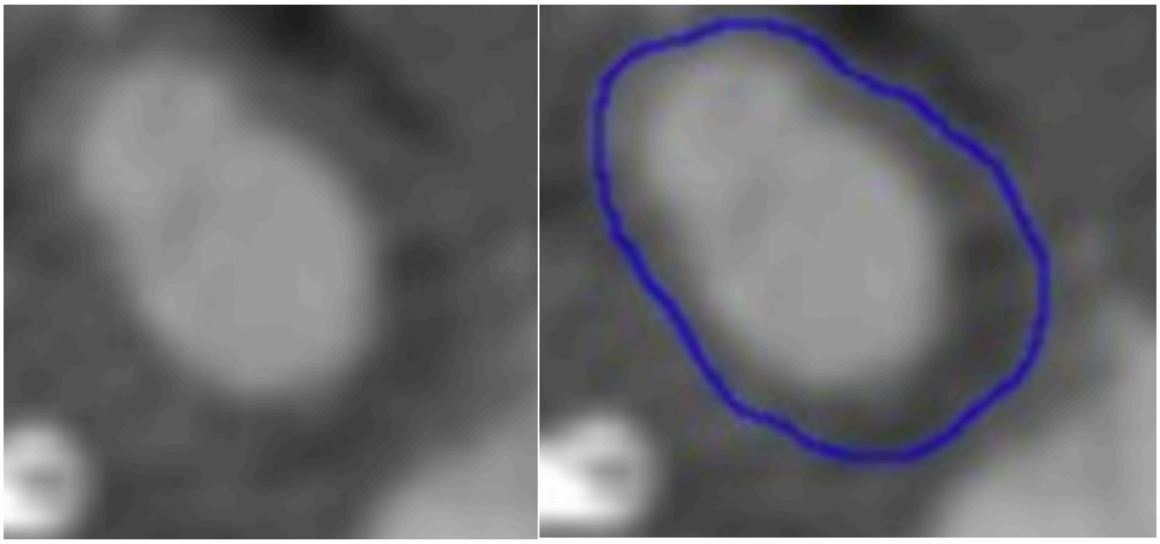


Figure 2.31 Axial CTA of the carotid artery bifurcation

This is an example of a slice identified as the carotid bifurcation (left image). An ROI was manually drawn in TexRAD to encompass the outer vessel wall and is shown in blue (right image).

2.3.6 TexRAD: Single-slice Approach

2.3.6.1 Symptomatic Patients: Culprit versus Non-culprit Carotids
2.3.6.1.1 Unfiltered Texture Features

A total of 82 carotids from 41 symptomatic patients: 41 culprit versus 41 non-culprit carotid arteries were included in this paired analysis. For single-slice analysis, the CT angiograms with a slice thickness of 0.625 mm were analysed. 4 algorithms were investigated in TexRAD that considered different ranges of Hounsfield units in the derivation of the texture features: (1) unrestricted algorithm considered a range of -1000 to 1000 HU, (2) restricted I algorithm considered a range of 0 to 200 HU, (3) restricted II algorithm considered a range from -1000 to 200 HU and (4) restricted III algorithm considered a range from -1000 to 300 HU.

Table 2.20 provides summary statistics of the histogram-derived texture features for culprit versus non-culprit carotid arteries using a single-slice approach (comparing carotid bifurcations) derived using each algorithm and without prior image filtration.

For each algorithm, 6 texture features were derived: (1) ‘mean’, (2) ‘mean of positive pixels’ (MPP), (3) ‘standard deviation’, (4) ‘entropy’, (5) ‘skewness’ and (6) ‘kurtosis’. However, please note in the restricted I algorithm, the TexRAD texture feature ‘mean of positive pixels’ is numerically equivalent to the texture feature ‘mean’ and so was discarded from analysis to reduce redundancy of information.

Table 2.20 TexRAD single-slice analysis: Unfiltered Features

Texture Feature	Culprit Artery – median (IQR)	Non-culprit Artery – median (IQR)	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	228.8 (160.1, 271.3)	241.9 (192.2, 316.5)	0.124
Mean of Positive Pixels	230.0 (161.1, 275.9)	248.5 (202.3, 320.3)	0.061
Standard Deviation	157.5 (116.4, 204.0)	165.8 (107.7, 203.4)	0.960
Entropy	5.27 (5.04, 5.37)	5.29 (5.10, 5.41)	0.119
Skewness	0.76 (0.31, 1.15)	0.78 (-0.04, 1.25)	0.412
Kurtosis	0.07 (-0.81, 0.95)	-0.07 (-1.24, 1.93)	0.562
Restricted I Algorithm [0, 200 HU]			
Mean	98.0 (86.7, 105.6)	101.4 (94.7, 114.2)	0.024*
Standard Deviation	46.4 (43.4, 49.6)	47.8 (45.2, 50.3)	0.067
Entropy	4.54 (4.33, 4.68)	4.55 (4.35, 4.69)	0.928
Skewness	0.33 (0.14, 0.53)	0.21 (-0.06, 0.42)	0.032*
Kurtosis	-0.63 (-0.89, -0.38)	-0.72 (-0.99, -0.49)	0.082
Restricted II Algorithm [-1000, 200 HU]			
Mean	94.7 (85.9, 104.0)	98.5 (89.3, 111.9)	0.258
Mean of Positive Pixels	98.0 (86.7, 105.6)	101.4 (94.7, 114.2)	0.024*
Standard Deviation	47.9 (44.5, 52.4)	50.2 (46.9, 54.9)	0.002*
Entropy	4.56 (4.33, 4.69)	4.60 (4.40, 4.74)	0.412
Skewness	0.24 (-0.02, 0.44)	-0.02 (-0.21, 0.20)	0.005*
Kurtosis	-0.42 (-0.75, -0.05)	-0.49 (-0.81, -0.24)	0.142
Restricted III Algorithm [-1000, 300 HU]			
Mean	134.5 (119.9, 141.9)	141.4 (122.5, 161.7)	0.069
Mean of Positive Pixels	136.0 (126.3, 143.4)	144.9 (126.2, 167.3)	0.018*
Standard Deviation	78.9 (70.3, 85.3)	78.7 (68.7, 88.1)	0.051
Entropy	4.88 (4.65, 4.99)	4.85 (4.68, 5.07)	0.230
Skewness	0.42 (0.14, 0.57)	0.18 (-0.12, 0.49)	0.001*
Kurtosis	-0.86 (-1.00, -0.54)	-0.77 (-1.08, -0.60)	0.960

*HU, Hounsfield units; IQR, interquartile range. Median and IQR reported. The p-value was obtained using Wilcoxon signed-rank test; *p-value<0.05.*

As shown in Table 2.20, in single-slice analysis, there were no differences between any TexRAD texture features of culprit versus non-culprit carotid arteries in symptomatic patients using the unrestricted algorithm and no prior image filtration. By restricting the range of Hounsfield units considered in the texture feature calculation (restricted algorithms I-III), some differences were identified between CC and NC carotid arteries.

- **Restricted I algorithm [0, 200 HU]**: compared with non-culprit carotid arteries, the culprit carotid had a lower 'mean' (p=0.024) and a more positive 'skewness' (p=0.032).
- **Restricted II algorithm [-1000, 200 HU]:** culprit carotid arteries had a lower 'MPP' (this is the same as the 'mean' extracted with the restricted I algorithm, p=0.024), a lower 'standard deviation' (p=0.002) and a more positive 'skewness' (p=0.005) than non-culprit carotid arteries.
- **Restricted III algorithm [-1000, 300 HU]:** consistent with the findings of the previous restricted algorithms, the culprit carotid was associated with a lower 'MPP' (p=0.018), and a more positive 'skewness' (p=0.001); there was no longer any difference in 'standard deviation' however.

To assess whether these differences in first order texture features were related to differences in carotid stenosis (the culprit carotid artery is associated with a greater degree of carotid stenosis than non-culprit carotid arteries), Spearman rank correlations were calculated between the TexRAD texture feature 'mean', which summarises the image histogram of the ROIs drawn (and from which the other TexRAD texture features are calculated) and the degree of carotid stenosis for culprit and non-culprit carotid arteries. In addition, Spearman rank correlations were calculated between the texture feature 'mean' and the level of carotid calcification to assess whether the restricted algorithms limited the effect of calcium upon the extracted texture features. Spearman's rho and associated p-values for the correlations are shown in Table 2.21 and the scatter plots between the respective variables are shown.

Table 2.21 Spearman's Rho correlation with Carotid Stenosis

Texture Feature	Correlation with	Spearman's rho and significance
Unrestricted Algorithm [-1000, 1000 HU]		
Mean	Carotid stenosis	r_s =-0.078, p=0.487
	Calcium	r_s =0.612, p<0.0001*
Restricted I Algorithm [0, 200 HU]		
Mean	Carotid stenosis	r_s =-0.183, p=0.100
	Calcium	r_s = 0.159, p=0.155
Restricted II Algorithm [-1000, 200 HU]		
Mean	Carotid stenosis	r_s =-0.122, p=0.273
	Calcium	r_s =0.148, p=0.185
Restricted III Algorithm [-1000, 300 HU]		
Mean	Carotid stenosis	r_s =0.231, p=0.037*
	Calcium	r_s =-0.003, p=0.976

*r_s, Spearman's rho; HU, Hounsfield units. *p-value<0.05.*

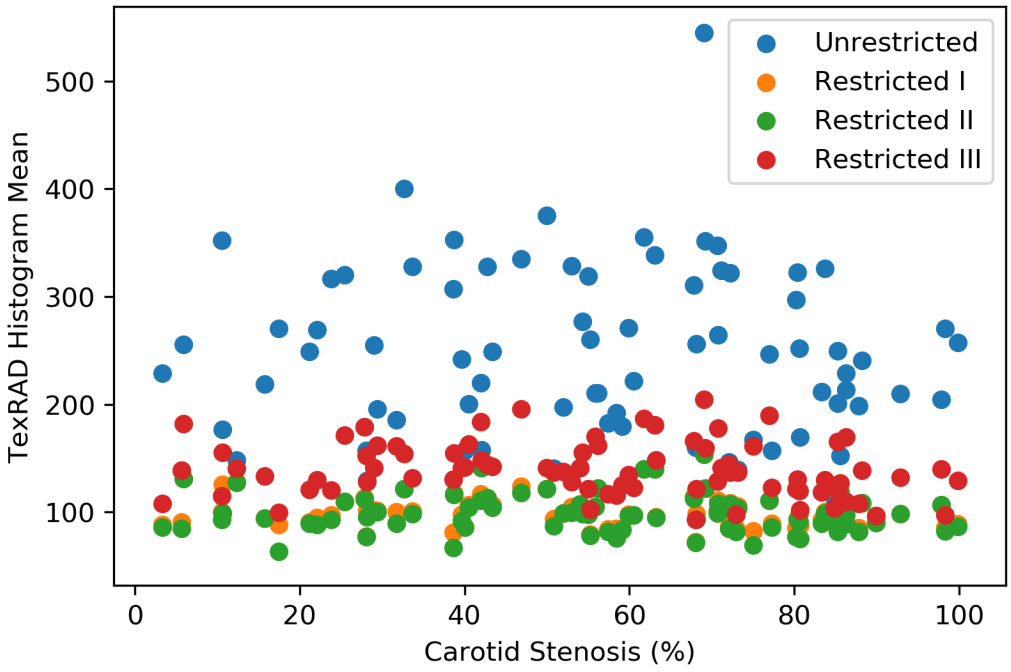


Figure 2.32 Scatter plot of carotid stenosis and TexRAD texture feature 'Mean'

As shown in Table 2.21 and Figure 2.32, there was no correlation between the TexRAD texture feature 'mean' and carotid stenosis, regardless of the algorithm used. There was an exception with restricted III algorithm [-1000, 300 HU] which had a weak positive correlation with carotid stenosis (r_s = *0.23, p*=0.037). Therefore, the TexRAD first order texture features were independent of carotid stenosis when using the unrestricted, restricted I or restricted II algorithms.

Figure 2.33 demonstrates a positive correlation between the TexRAD texture feature 'mean' and the carotid calcium burden (r_s =0.61, p<0.0001) when using the unrestricted algorithm. This was no longer the case when using the restricted algorithms (p>0.05).

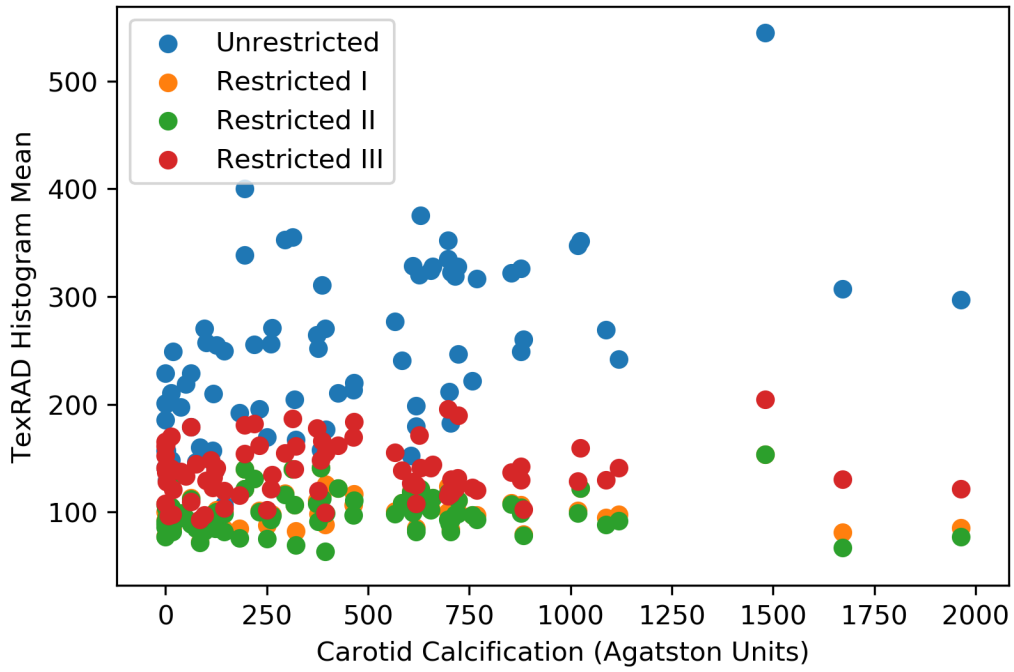


Figure 2.33 Scatter plot of Carotid Calcification and TexRAD texture feature 'Mean'

2.3.6.1.2 Filtered Texture Features: Laplacian of Gaussian Filter

The effect of image filtering using a LoG filter of sizes 2-6 mm prior to texture feature calculation was investigated because these settings are frequently used in TexRAD studies[169]. Table 2.22 gives the median and IQR of the TexRAD texture features in culprit and non-culprit carotid arteries when the LoG filter (SSF=2 mm) was applied to the image in single-slice analysis.

Following prior filtering, there were no differences in texture features between culprit and non-culprit carotid arteries, regardless of algorithm used. Please note, the LoG filter changes the meaning of the grey level values such that they are no longer calibrated as Hounsfield units, therefore in the restricted I algorithm, the texture feature 'mean' and the texture feature 'mpp' are no longer equivalent.

There were also no differences between culprit versus non-culprit carotid arteries when using larger LoG filters (SSF 3-6 mm) when using the restricted algorithms (I-III). When using the unrestricted algorithm, there were 3 features that had a p-value <0.05, these are listed in Table 2.23.

Table 2.22 TexRAD single-slice analysis - 0.625 mm slice thickness: LoG filter (SSF = 2)

Texture Feature	Culprit Artery – median (IQR)	Non-culprit Artery – median (IQR)	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	115.7 (58.0, 191.3)	161.5 (126.2, 214.7)	0.052
Mean of Positive Pixels	441.7 (342.0, 624.1)	485.7 (353.9, 660.9)	0.976
Standard Deviation	484.5 (346.9, 693.5)	518.5 (318.7, 653.9)	0.834
Entropy	5.62 (5.51, 5.82)	5.69 (5.49, 5.91)	0.066
Skewness	0.46 (0.09, 0.94)	0.61 (-0.02, 1.12)	1.000
Kurtosis	-0.08 (-0.75, 1.36)	0.11 (-1.01, 1.14)	0.516
Restricted I Algorithm [0, 200 HU]			
Mean	-179.5 (-283.6, -105.4)	-149.4 (-232.3, -77.3)	0.061
Mean of Positive Pixels	147.1 (109.6, 159.8)	160.1 (122.2, 199.8)	0.124
Standard Deviation	236.5 (188.6, 314.5)	243.2 (181.0, 289.4)	0.276
Entropy	4.95 (4.69, 5.19)	4.91 (4.66, 5.23)	0.418
Skewness	0.17 (-0.08, 0.44)	0.34 (0.04, 0.63)	0.465
Kurtosis	0.00 (-0.38, -0.79)	-0.20 (-0.49, 0.81)	0.280
Restricted II Algorithm [-1000, 200 HU]			
Mean	-192.6 (-283.6, -104.8)	-151.2 (-232.4, -96.8)	0.089
Mean of Positive Pixels	147.1 (109.6, 159.8)	160.1 (122.2, 198.5)	0.131
Standard Deviation	236.5 (188.8, 316.7)	248.2 (181.8, 298.4)	0.384

Entropy	4.95 (4.71, 5.21)	4.96 (4.71, 5.27)	0.610
Skewness	0.17 (-0.05, 0.45)	0.30 (0.03, 0.59)	0.503
Kurtosis	0.07 (-0.38, 0.83)	-0.18 (-0.49, 0.83)	0.294
Restricted III Algorithm [-1000, 300 HU]			
Mean	-109.0 (-193.3, 17.5)	-45.8 (-171.1, 46.2)	0.215
Mean of Positive Pixels	233.3 (201.8, 269.6)	252.2 (219.9, 275.1)	0.134
Standard Deviation	296.9 (250.2, 360.7)	314.5 (262.0, 354.5)	0.865
Entropy	5.25 (5.0, 5.5)	5.27 (5.04, 5.56)	0.849
Skewness	0.23 (0.04, 0.48)	0.27 (0.03, 0.56)	0.582
Kurtosis	-0.14 (-0.58, 0.57)	-0.15 (-0.69, 0.77)	0.841

*HU, Hounsfield units; IQR, interquartile range. Median and IQR reported. The p-value was obtained using Wilcoxon signed-rank test; *p-value<0.05.*

Comparative analysis of the filtered features involved 6 different filter sizes (2-6 mm) and 6 TexRAD texture features for each setting. Therefore, for each algorithm, 36 texture features were extracted from the ROI. If we apply the Benjamini-Hochberg correction for multiple testing, which is less conservative than Bonferroni correction, then these differences are no longer statistically significant. In addition, considering the lack of consistency in differences of these texture features (shown in Table 2.22 and Table 2.23) in the other SSF sizes and algorithms, these differences likely reflect false positive results from multiple comparisons. It was therefore decided to continue texture analysis investigation using unfiltered texture features only for the remainder of the study.

Table 2.23 Unrestricted algorithm Texture Features with LoG Filter

Texture Feature	**SSF**	**Culprit Artery – median (IQR)**	**Non-culprit Artery – median (IQR)**	***p*-value**
Unrestricted Algorithm [-1000, 1000 HU]				
Entropy	3	5.62 (5.51, 5.89)	5.71 (5.52, 5.93)	0.026*
Entropy	4	5.62 (5.51, 5.84)	5.73 (5.51, 5.93)	0.027*
Skewness	6	0.07 (-0.26, 0.29)	-0.11 (-0.27, 0.07)	0.009*

*SSF, spatial scale filters; IQR, interquartile range. *p-value<0.05.*

2.3.6.2 Inter-patient comparisons: Asymptomatic versus Symptomatic
2.3.6.2.1 Carotid Artery Analysis

Differences in TexRAD's first order texture features between asymptomatic and symptomatic patients were evaluated on a per-vessel basis i.e. asymptomatic versus culprit carotid arteries and asymptomatic versus non-culprit carotid arteries. There were 50 asymptomatic carotid arteries, 41 culprit and 41 non-culprit carotid arteries included in this analysis. Please note that not all asymptomatic carotid arteries in this analysis had carotid atheroma (22 out of the 50 asymptomatic carotid arteries had no plaque, see Table 2.17).

With the unrestricted algorithm, there were differences in all features between the carotid arteries of asymptomatic patients, and those of symptomatic patients, except for the texture feature 'standard deviation'. Asymptomatic carotid arteries had a higher 'mean' (269.4) than culprit (228.8) and non-culprit (241.9) carotids. Similarly, Asx carotids had a higher 'mean of positive pixels' (272.5), than CC (230.0) and NC (248.5). Asx carotids were associated with a lower 'entropy' (5.12) compared with CC (5.27) and NC (5.29) carotid arteries; a more negative 'skewness' and a more negative 'kurtosis'.

Restricting the range of Hounsfield units with the restricted algorithms (I-III), decreased the detectable differences between asymptomatic carotid arteries and symptomatic carotid arteries, shown in Table 2.24. The consistent differences, regardless of algorithm were: (1) asymptomatic carotid arteries had a higher 'mean' than culprit carotid arteries, (2) asymptomatic carotid arteries had a more negative 'skewness' than culprit carotid arteries and (3) asymptomatic carotid arteries had a lower 'entropy' than non-culprit carotid arteries.

Table 2.24 TexRAD single-slice analysis: Asx vs CC or NC

Texture Feature	Asymptomatic Artery – median (IQR)	Comparisons	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	269.4 (228.8, 313.0)	Asx vs CC	0.004*
		Asx vs NC	0.034*
MPP	272.5 (232.4, 316.8)	Asx vs CC	0.003*
		Asx vs NC	0.034*
Standard Deviation	142.0 (109.4, 167.8)	Asx vs CC	0.219
		Asx vs NC	0.178
Entropy	5.12 (5.02, 5.24)	Asx vs CC	0.030*
		Asx vs NC	0.006*
Skewness	-0.48 (-0.77, 0.01)	Asx vs CC	<0.0001*

		Asx vs NC	<0.0001*
Kurtosis	-0.91 (-1.28, -0.36)	Asx vs CC	0.001*
		Asx vs NC	0.006*
Restricted I Algorithm [0, 200 HU]			
Mean	107.1 (97.1, 114.1)	Asx vs CC	0.002*
		Asx vs NC	0.542
Standard Deviation	48.2 (45.0, 50.4)	Asx vs CC	0.100
		Asx vs NC	0.780
Entropy	4.37 (4.23, 4.53)	Asx vs CC	0.006*
		Asx vs NC	0.001*
Skewness	0.16 (-0.10, 0.24)	Asx vs CC	0.001*
		Asx vs NC	0.315
Kurtosis	-0.84 (-1.01, -0.69)	Asx vs CC	0.002*
		Asx vs NC	0.105
Restricted II Algorithm [-1000, 200 HU]			
Mean	101.7 (91.0, 111.0)	Asx vs CC	0.037*
		Asx vs NC	0.649
MPP	107.1 (97.1, 114.1)	Asx vs CC	0.002*
		Asx vs NC	0.536
Standard Deviation	51.9 (48.4, 55.7)	Asx vs CC	0.004*
		Asx vs NC	0.599
Entropy	4.40 (4.28, 4.55)	Asx vs CC	0.018*
		Asx vs NC	0.002*
Skewness	-0.05 (-0.31, 0.13)	Asx vs CC	<0.0001*
		Asx vs NC	0.468
Kurtosis	-0.55 (-0.80, -0.36)	Asx vs CC	0.149
		Asx vs NC	0.814
Restricted III Algorithm [-1000, 300 HU]			
Mean	148.7 (137.3, 163.7)	Asx vs CC	0.001*
		Asx vs NC	0.173
MPP	155.3 (140.0, 168.1)	Asx vs CC	<0.0001*
		Asx vs NC	0.156
Standard Deviation	82.4 (78.7, 87.2)	Asx vs CC	0.016*
		Asx vs NC	0.120
Entropy	4.72 (4.61, 4.91)	Asx vs CC	0.102
		Asx vs NC	0.013*
Skewness	0.17 (-0.11, 0.41)	Asx vs CC	0.002*
		Asx vs NC	0.681
Kurtosis	-0.86 (-1.08, -0.62)	Asx vs CC	0.585
		Asx vs NC	0.672

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; HU, Hounsfield units; MPP, mean of positive pixels. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

2.3.6.2.2 Comparisons with asymptomatic carotid plaque only

The analysis was repeated comparing the carotid arteries from asymptomatic patients with those from symptomatic patients, however restricted to only carotid arteries with plaque (28 asymptomatic carotid arteries). With the unrestricted algorithm, when restricting the analysis to comparisons of asymptomatic carotid arteries with plaque only and culprit or non-culprit carotid arteries respectively, 'entropy' was no longer statistically significantly different. 'Kurtosis' was no longer statistically significantly different between asymptomatic and non-culprit carotid arteries. The remaining texture features followed the same pattern of differences described in section 2.3.6.2.1.

Using the restricted algorithms, we see a similar pattern of differences between asymptomatic and symptomatic carotid arteries whether or not we only compare asymptomatic carotid arteries with carotid plaque. However, with the restricted II algorithm, the texture feature 'mean' is no longer statistically significantly different between Asx versus CC carotid arteries, nor is 'entropy'.

Table 2.25 TexRAD single-slice analysis: Asx (plaque only) vs CC or NC

Texture Feature	Asymptomatic Artery – median (IQR)	Comparisons	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	285.1 (242.1, 324.5)	Asx vs CC	0.005*
		Asx vs NC	0.023*
MPP	291.1 (246.3, 328.5)	Asx vs CC	0.004*
		Asx vs NC	0.024*
Standard Deviation	158.4 (113.9, 197.3)	Asx vs CC	0.881
		Asx vs NC	0.889
Entropy	5.16 (5.09, 5.32)	Asx vs CC	0.478
		Asx vs NC	0.197
Skewness	-0.120 (-0.593, 0.280)	Asx vs CC	<0.0001*
		Asx vs NC	0.0003*
Kurtosis	-0.785 (-1.20, 0.028)	Asx vs CC	0.021*
		Asx vs NC	0.077
Restricted I Algorithm [0, 200 HU]			
Mean	101.8 (95.7, 112.3)	Asx vs CC	0.049*
		Asx vs NC	0.873
Standard Deviation	47.8 (44.8, 50.5)	Asx vs CC	0.238
		Asx vs NC	0.984
Entropy	4.36 (4.15, 4.53)	Asx vs CC	0.023*
		Asx vs NC	0.009*
Skewness	0.185 (0.048, 0.265)	Asx vs CC	0.022*

		Asx vs NC	0.803
Kurtosis	-0.790 (-1.03, -0.650)	Asx vs CC	0.013*
		Asx vs NC	0.234
Restricted II Algorithm [-1000, 200 HU]			
Mean	99.9 (90.3, 105.2)	Asx vs CC	0.352
		Asx vs NC	0.726
MPP	101.8 (95.7, 112.3)	Asx vs CC	0.049*
		Asx vs NC	0.873
Standard Deviation	51.6 (48.1, 55.2)	Asx vs CC	0.030*
		Asx vs NC	0.841
Entropy	4.40 (4.22, 4.57)	Asx vs CC	0.061
		Asx vs NC	0.014*
Skewness	0.03 (-0.156, 0.163)	Asx vs CC	0.009*
		Asx vs NC	0.873
Kurtosis	-0.625 (-0.805, -0.333)	Asx vs CC	0.226
		Asx vs NC	0.841
Restricted III Algorithm [-1000, 300 HU]			
Mean	145.6 (132.4, 160.4)	Asx vs CC	0.036*
		Asx vs NC	0.689
MPP	149.7 (138.5, 165.6)	Asx vs CC	0.015*
		Asx vs NC	0.603
Standard Deviation	84.0 (74.7, 91.1)	Asx vs CC	0.023*
		Asx vs NC	0.171
Entropy	4.73 (4.55, 4.91)	Asx vs CC	0.180
		Asx vs NC	0.039*
Skewness	0.260 (-0.103, 0.450)	Asx vs CC	0.035*
		Asx vs NC	0.841
Kurtosis	-0.955 (-1.17, -0.490)	Asx vs CC	0.390
		Asx vs NC	0.478

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; HU, Hounsfield units; MPP, mean of positive pixels. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

2.3.7 TexRAD Multi-slice Analysis

2.3.7.1 Symptomatic Patients: Culprit versus Non-culprit Carotids
2.3.7.1.1 Cumulative Histogram Texture Parameters

In multi-slice analysis, ROIs were drawn on 14 consecutive slices of the carotid artery, reporting on the common carotid artery and internal carotid artery. Previously in the single-slice approach, there were no differences between culprit versus non-culprit carotid arteries when using the unrestricted algorithm. However, in the multi-slice approach, even with the unrestricted algorithm, there were differences in first order features between CC and NC carotid arteries. In this algorithm, culprit carotid arteries had a lower 'mean', a lower 'mpp', lower 'entropy' and a more positive 'skewness' than non-culprit carotid arteries.

With the restricted algorithms (I-III), there were differences in all texture features between CC and NC carotid arteries that followed a similar relationship: culprit carotids were associated with a lower 'mean', a lower 'standard deviation', a lower 'entropy', a more positive 'skewness' and a less negative 'kurtosis' than non-culprit carotid arteries, see Table 2.26.

Table 2.26 Multi-slice Cumulative Histogram Features: Culprit versus Non-culprit

Texture Features	Culprit Artery – median (IQR)	Non-culprit Artery – median (IQR)	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	136.1 (89.6, 226.5)	241.9 (192.2, 316.5)	<0.0001*
MPP	136.1 (92.2, 252.1)	248.5 (202.3, 320.3)	<0.0001*
Standard Deviation	106.5 (58.8, 217.1)	165.8 (107.7, 203.4)	0.060
Entropy	4.91 (4.57, 5.07)	5.29 (5.1, 5.41)	<0.0001*
Skewness	1.21 (0.68, 1.58)	0.78 (-0.04, 1.25)	0.038*
Kurtosis	1.29 (0.12, 3.34)	-0.07 (-1.24, 1.93)	0.082
Restricted I Algorithm [0, 200 HU]			
Mean	95.1 (75.8, 105.9)	101.4 (94.7, 114.2)	<0.0001*
Standard Deviation	43.4 (38.4, 47.6)	47.8 (45.2, 50.3)	0.002*
Entropy	4.38 (4.17, 4.58)	4.55 (4.35, 4.69)	0.012*
Skewness	0.45 (0.18, 0.64)	0.21 (-0.06, 0.42)	0.005*
Kurtosis	-0.53 (-0.84, 0.13)	-0.72 (-0.99, -0.49)	0.008*
Restricted II Algorithm [-1000, 200 HU]			
Mean	93.6 (71.1, 104.2)	98.5 (89.3, 111.9)	<0.0001*
MPP	95.1 (75.8, 105.9)	101.4 (94.7, 114.2)	<0.0001*
Standard Deviation	45.7 (38.7, 50.5)	50.2 (46.9, 54.9)	0.001*
Entropy	4.45 (4.26, 4.60)	4.60 (4.40, 4.74)	0.013*
Skewness	0.19 (-0.07, 0.52)	-0.02 (-0.21, 0.20)	0.014*
Kurtosis	-0.24 (-0.73, 0.34)	-0.49 (-0.81, -0.24)	0.014*

Restricted III Algorithm [-1000, 300 HU]			
Mean	110.1 (77.6, 131.7)	141.4 (122.5, 161.7)	<0.0001*
MPP	111.0 (82.9, 134.6)	144.9 (126.2, 167.3)	<0.0001*
Standard Deviation	66.4 (53.9, 75.9)	78.7 (68.7, 88.1)	<0.0001*
Entropy	4.58 (4.41, 4.78)	4.85 (4.68, 5.07)	<0.0001*
Skewness	0.73 (0.17, 1.07)	0.18 (-0.12, 0.49)	<0.0001*
Kurtosis	0.12 (-0.73, 1.17)	-0.77 (-1.08, -0.6)	<0.0001*

*MPP, mean of positive pixels; SD, standard deviation; IQR, interquartile range. The P value was obtained using the Wilcoxon signed-rank test, after assessing for normality of differences or a paired t-test as appropriate. * p<0.05.*

2.3.7.1.2 Carotid Plaque Type: Subset Analysis

When comparing culprit versus non-culprit carotid arteries in symptomatic patients, paired analysis is required. For subset analysis of the TexRad texture features according to carotid plaque type in symptomatic patients (intra-patient comparisons): (1) calcified, (2) noncalcified and (3) mixed, there were 8 pairs of carotid arteries with calcified carotid plaque, 2 noncalcified pairs and 20 mixed pairs. Since the sample size for the noncalcified carotid plaque was small, interpretation of that subset analysis was limited.

Previously in multi-slice analysis, when comparing culprit with non-culprit carotid arteries regardless of plaque type, there were statistically significant differences in all TexRAD texture features except ‘standard deviation’. In subset analysis, there were no longer differences in ‘mean’, ‘mean of positive pixels’, ‘skewness’ or ‘kurtosis’ when comparing paired calcified carotid arteries in symptomatic patients and when comparing paired noncalcified carotid arteries.

TexRAD feature differences were found mostly in the carotid arteries with the mixed plaque type: (1) culprit carotid arteries with mixed plaque had a lower ‘mean’ than non-culprit carotid arteries with mixed plaque (p=0.002), (2) culprit carotid arteries had a lower ‘mean of positive pixels’ than non-culprit carotids (p=0.003) and (3) culprit carotid arteries with mixed plaque had a lower ‘entropy’ than non-culprit carotid arteries with mixed plaque (p<0.0001).

Following resegmentation with the restricted algorithms I-III, there were still no differences between culprit versus non-culprit carotid arteries with calcified plaques nor between those with noncalcified plaques. In resegmentation, the general pattern for culprit carotid arteries with mixed plaque versus those of non-culprit carotid arteries was as follows: (1) culprit carotid

arteries had a lower 'mean' than non-culprit carotid arteries, (2) culprit carotids had a lower 'mpp', (3) culprit carotids had a lower 'standard deviation', (4) culprit carotids had a lower 'entropy', (5) culprit carotids had a more positive 'skewness' and (6) culprit carotid arteries had a more positive 'kurtosis' than non-culprit carotid arteries with mixed plaque.

Table 2.27 Multi-slice Carotid Plaque Type Subset Analysis: Culprit versus Non-culprit

Texture Feature	Plaque Type	Culprit Artery – median (IQR)	Non-culprit Artery – median (IQR)	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]				
Mean	Calcified	337.8 (130.5, 392.2)	312.9 (259.5, 327.9)	0.889
	Noncalcified	76.6 (76.3, 76.9)	149.1 (145.1, 153.1)	0.180
	Mixed	108.9 (84.3, 196.1)	210.4 (171.4, 247.1)	0.002*
MPP	Calcified	343.2 (131.4, 392.5)	321.3 (265.5, 329.4)	0.889
	Noncalcified	76.6 (76.3, 76.9)	149.6 (145.3, 153.8)	0.180
	Mixed	109.8 (90.6, 223.9)	210.4 (181.2, 249.9)	0.003*
Standard Deviation	Calcified	237.7 (74.4, 265.5)	198.0 (179.7, 218.9)	0.674
	Noncalcified	31.0 (26.4, 35.5)	85.4 (75.7, 95.2)	0.180
	Mixed	100.7 (60.2, 200.2)	144.7 (88.6, 174.2)	0.093
Entropy	Calcified	5.15 (4.68, 5.37)	5.39 (5.33, 5.53)	0.028*
	Noncalcified	4.36 (4.25, 4.46)	4.99 (4.91, 5.06)	0.180
	Mixed	4.90 (4.77, 5.01)	5.25 (5.04, 5.41)	<0.0001*
Skewness	Calcified	0.775 (0.535, 1.48)	0.530 (-0.203, 1.02)	1.000
	Noncalcified	0.585 (0.353, 0.818)	0.100 (0.100, 0.100)	0.180
	Mixed	1.26 (0.763, 1.69)	1.10 (-0.105, 1.70)	0.332
Kurtosis	Calcified	-0.300 (-0.800, 3.32)	-0.085 (-1.188, 0.990)	0.779
	Noncalcified	1.04 (0.58, 1.50)	-1.50 (-1.57, -1.44)	0.180
	Mixed	1.26 (0.618, 3.75)	0.930 (-0.993, 6.33)	0.723
Restricted I Algorithm [0, 200 HU]				
Mean	Calcified	104.3 (88.9, 116.4)	103.7 (91.2, 121.1)	0.674
	Noncalcified	75.3 (75.0, 75.7)	92.5 (85.2, 99.9)	0.180
	Mixed	82.2 (73.8, 95.3)	103.2 (91.6, 116.0)	<0.0001*
Standard Deviation	Calcified	45.4 (34.1, 52.1)	47.5 (43.7, 48.1)	0.779
	Noncalcified	28.7 (25.3, 32.1)	47.2 (46.2, 48.3)	0.180
	Mixed	44.8 (39.4, 49.2)	48.8 (47.1, 52.1)	0.012*
Entropy	Calcified	4.20 (3.90, 4.35)	4.32 (4.28, 4.73)	0.123
	Noncalcified	4.34 (4.25, 4.44)	4.62 (4.57, 4.66)	0.317
	Mixed	4.50 (4.32, 4.61)	4.62 (4.46, 4.69)	0.100
Skewness	Calcified	0.175 (-0.160, 0.525)	0.165 (-0.060, 0.425)	0.889
	Noncalcified	0.320 (0.22, 0.42)	0.520 (0.405, 0.635)	0.655
	Mixed	0.620 (0.293, 0.850)	0.275 (-0.098, 0.450)	0.002*
Kurtosis	Calcified	-0.935 (-1.07, 0.550)	-0.625 (-0.873, -0.420)	0.674
	Noncalcified	0.195 (0.158, 0.233)	-0.565 (-0.783, -0.348)	0.180
	Mixed	-0.275 (-0.813, 0.755)	-0.790 (-1.03, -0.498)	0.005*
Restricted II Algorithm [-1000, 200 HU]				
Mean	Calcified	100.4 (87.4, 115.4)	101.5 (74.8, 116.7)	0.889

	Noncalcified	75.3 (75.0, 75.7)	92.1 (84.6, 99.7)	0.180
	Mixed	74.2 (67.3, 93.7)	98.0 (89.9, 111.6)	0.001*
MPP	Calcified	104.3 (88.9, 116.4)	103.7 (91.2, 121.1)	0.674
	Noncalcified	75.3 (75.0, 75.7)	92.5 (85.2, 99.9)	0.180
	Mixed	82.2 (73.8, 95.3)	103.2 (91.6, 116.0)	<0.0001*
Standard Deviation	Calcified	45.8 (35.0, 54.5)	48.7 (45.3, 58.3)	0.327
	Noncalcified	28.7 (25.3, 32.1)	47.5 (46.3, 48.6)	0.180
	Mixed	45.9 (41.7, 52.1)	52.6 (50.0, 56.8)	0.044*
Entropy	Calcified	4.29 (3.91, 4.38)	4.42 (4.29, 4.74)	0.123
	Noncalcified	4.34 (4.25, 4.44)	4.63 (4.58, 4.67)	0.317
	Mixed	4.57 (4.40, 4.65)	4.69 (4.54, 4.76)	0.135
Skewness	Calcified	-0.080 (-0.273, 0.258)	-0.045 (-0.298, 0.120)	0.674
	Noncalcified	0.320 (0.22, 0.42)	0.500 (0.395, 0.605)	0.655
	Mixed	0.465 (0.098, 0.640)	-0.005 (-0.210, 0.208)	0.008*
Kurtosis	Calcified	-0.700 (-1.02, 0.773)	-0.395 (-0.808, -0.063)	0.674
	Noncalcified	0.195 (0.158, 0.233)	-0.565 (-0.783, -0.348)	0.180
	Mixed	-0.110 (-0.805, 0.760)	-0.605 (-0.863, -0.278)	0.013*
Restricted III Algorithm [-1000, 300 HU]				
Mean	Calcified	129.5 (107.0, 179.0)	137.0 (123.9, 172.3)	0.401
	Noncalcified	76.6 (76.3, 76.9)	146.8 (144.0, 149.7)	0.180
	Mixed	90.6 (72.9, 113.2)	140.9 (122.7, 164.8)	<0.0001*
MPP	Calcified	133.3 (108.1, 179.0)	144.4 (127.3, 172.8)	0.484
	Noncalcified	76.6 (76.3, 76.9)	147.2 (144.2, 150.3)	0.180
	Mixed	102.5 (82.5, 118.6)	146.1 (127.6, 169.3)	<0.0001*
Standard Deviation	Calcified	74.8 (47.2, 81.9)	77.5 (68.5, 94.6)	0.161
	Noncalcified	31.0 (26.4, 35.5)	84.6 (75.3, 93.9)	0.180
	Mixed	69.1 (55.1, 73.9)	80.2 (75.5, 89.7)	0.009*
Entropy	Calcified	4.57 (4.25, 4.62)	4.82 (4.57, 5.11)	0.050
	Noncalcified	4.36 (4.25, 4.46)	4.97 (4.90, 5.03)	0.180
	Mixed	4.76 (4.50, 4.86)	4.98 (4.83, 5.06)	0.002*
Skewness	Calcified	0.290 (0.013, 0.893)	0.315 (-0.380, 0.558)	0.575
	Noncalcified	0.585 (0.353, 0.818)	0.125 (0.113, 0.138)	0.655
	Mixed	0.825 (0.408, 1.14)	0.055 (-0.263, -0550)	0.001*
Kurtosis	Calcified	-0.840 (-1.08, 1.45)	-0.700 (-0.758, -0.635)	0.575
	Noncalcified	1.04 (0.58, 1.50)	-1.49 (-1.55, -1.43)	0.180
	Mixed	0.395 (-0.315, 1.13)	-.0.750 (-1.08, -0.433)	<0.0001*

*MPP, mean of positive pixels. Median and IQR reported. There were 8 pairs of carotid arteries for calcified carotid plaques, 2 pairs for noncalcified carotid plaques and 20 pairs for mixed carotid plaques. The p-value was obtained using Wilcoxon signed-rank test; *p-value<0.05.*

2.3.7.1.3 Carotid Stenosis Severity: Subset Analysis

In paired analysis, there were 3 symptomatic carotid pairs with carotid stenosis below 50% and 14 carotid pairs with ≥ 50% stenosis. The small sample size for the below 50% carotid stenosis category limited interpretation of the data and the identification of any differences between CC and NC carotid arteries.

For the ≥ 50% stenosis group, only 'entropy' was statistically significantly different between CC and NC carotid arteries with culprit carotid arteries being associated with a lower 'entropy' than non-culprit carotid arteries. Following resegmentation with restricted algorithms I and II, no differences could be found between culprit and non-culprit carotid arteries. There were differences using restricted III however for the carotid arteries with ≥ 50% stenosis: (1) culprit carotids had lower 'mean' than non-culprit carotid arteries, (2) culprit carotids had a lower 'mpp', (3) culprit carotids had a lower 'entropy' and (4) culprit carotids had a more positive 'kurtosis' than non-culprit carotid arteries. These findings were similar to the paired symptomatic carotid analysis in section 2.3.7.1.1.

Table 2.28 Multi-slice Carotid Stenosis Subset Analysis: Culprit versus Non-culprit

Texture Feature	Carotid Stenosis Severity	Culprit Artery – median (IQR)	Non-culprit Artery – median (IQR)	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]				
Mean	1-50%	195.3 (165.7, 203.8)	200.2 (193.0, 224.6)	0.109
	≥ 50%	160.6 (108.2, 276.6)	217.6 (196.8, 312.9)	0.056
MPP	1-50%	196.9 (166.5, 205.1)	202.3 (199.0, 226.7)	0.109
	≥ 50%	176.1 (109.0, 276.6)	221.3 (202.9, 313.9)	0.096
Standard Deviation	1-50%	126.6 (95.9, 147.9)	107.6 (104.8, 114.1)	0.593
	≥ 50%	154.8 (63.5, 219.3)	165.9 (122.0, 208.1)	0.397
Entropy	1-50%	5.02 (4.73, 5.10)	5.26 (5.20, 5.29)	0.109
	≥ 50%	4.88 (4.77, 5.03)	5.26 (5.13, 5.43)	0.001*
Skewness	1-50%	1.49 (1.19, 1.53)	-0.30 (-0.38, -0.26)	0.109
	≥ 50%	1.17 (0.86, 1.77)	1.07 (0.048, 1.26)	0.158
Kurtosis	1-50%	2.99 (1.45, 3.82)	-1.33 (-1.34, -1.29)	0.109
	≥ 50%	1.54 (0.54, 2.89)	0.885 (-1.11, 1.63)	0.140
Restricted I Algorithm [0, 200 HU]				
Mean	1-50%	104.2 (103.4, 108.3)	106.8 (103.6, 106.8)	1.000
	≥ 50%	95.1 (75.6, 108.7)	102.9 (94.7, 109.0)	0.074
Standard Deviation	1-50%	43.1 (41.8, 43.2)	46.9 (45.3, 47.0)	0.109
	≥ 50%	44.8 (41.8, 47.6)	47.7 (44.3, 50.0)	0.272
Entropy	1-50%	4.38 (4.28, 4.42)	4.60 (4.52, 4.63)	0.109
	≥ 50%	4.38 (3.99, 4.61)	4.50 (4.38, 4.60)	0.152

Skewness	1-50%	0.33 (0.30, 0.43)	0.12 (0.035, 0.19)	0.109
	≥ 50%	0.420 (0.080, 0.613)	0.33 (-0.008, 0.468)	0.245
Kurtosis	1-50%	-0.56 (-0.64, -0.56)	-0.51 (-0.55, -0.38)	0.109
	≥ 50%	-0.600 (-1.03, -0.053)	-0.670 (-1.01, -0.408)	0.272
Restricted II Algorithm [-1000, 200 HU]				
Mean	1-50%	102.4 (102.1, 107.4)	104.3 (96.8, 104.3)	1.000
	≥ 50%	94.1 (73.7, 108.7)	98.3 (90.1, 108.6)	0.084
MPP	1-50%	104.2 (103.4, 108.3)	106.8 (103.6, 106.8)	1.000
	≥ 50%	95.1 (75.6, 108.7)	102.9 (94.7, 109.0)	0.074
Standard Deviation	1-50%	43.9 (42.3, 45.1)	49.5 (48.0, 53.6)	0.109
	≥ 50%	45.8 (41.9, 47.6)	50.2 (45.1, 54.8)	0.158
Entropy	1-50%	4.39 (4.28, 4.44)	4.64 (4.59, 4.67)	0.109
	≥ 50%	4.48 (4.24, 4.61)	4.55 (4.39, 4.70)	0.116
Skewness	1-50%	0.19 (0.05, 0.355)	-0.19 (-0.25, -0.14)	0.109
	≥ 50%	0.335 (0.075, 0.543)	0.025 (-0.128, 0.313)	0.109
Kurtosis	1-50%	-0.48 (-0.595, -0.005)	-0.10 (-0.23, -0.075)	1.000
	≥ 50%	-0.435 (-1.01, 0.053)	-0.530 (-0.835, -0.263)	0.470
Restricted III Algorithm [-1000, 300 HU]				
Mean	1-50%	134.6 (131.0, 147.3)	161.5 (152.0, 162.3)	0.109
	≥ 50%	111.1 (76.1, 147.8)	134.5 (119.9, 166.8)	0.026*
MPP	1-50%	134.6 (131.5, 148.1)	165.5 (155.2, 168.7)	0.109
	≥ 50%	116.9 (88.7, 147.8)	137.1 (122.6, 167.9)	0.035*
Standard Deviation	1-50%	71.7 (67.6, 79.0)	87.6 (82.6, 93.7)	0.109
	≥ 50%	71.0 (58.3, 76.5)	78.5 (68.5, 89.6)	0.056
Entropy	1-50%	4.63 (4.52, 4.80)	5.03 (5.01, 5.05)	0.109
	≥ 50%	4.56 (4.35, 4.80)	4.85 (4.66, 5.04)	0.004*
Skewness	1-50%	0.84 (0.49, 0.85)	0.05 (-0.035, 0.235)	0.109
	≥ 50%	0.855 (0.125, 1.07)	0.275 (-0.073, 0.613)	0.079
Kurtosis	1-50%	-0.15 (-0.675, -0.09)	-1.20 (-1.21, -0.96)	0.180
	≥ 50%	0.215 (-0.955, 0.955)	-0.775 (-1.24, -0.313)	0.045*

*MPP, mean of positive pixels; HU, Hounsfield units; IQR, interquartile range. Median and IQR reported. For the group 1-50% carotid stenosis, there were 3 carotid pairs, for the group ≥ 50%, there were 14 pairs. The p-value was obtained using Wilcoxon signed-rank test; *p-value<0.05.*

2.3.7.2 Inter-patient comparisons: Asymptomatic versus Symptomatic
2.3.7.2.1 Carotid Artery Analysis

When considering all asymptomatic carotid arteries (n=50), including those without CT-defined carotid plaque, in the per-vessel comparison of asymptomatic patients versus symptomatic patients, there were some differences in texture features across the different algorithms. When using the unrestricted algorithm, asymptomatic carotids had a higher 'mean' than culprit carotids, but there was no difference with non-culprit carotid arteries. Asymptomatic carotid arteries had a *higher* 'entropy' compared with culprit and non-culprit carotid arteries in multi-slice analysis; this was opposite to the relationship seen when comparing only the carotid bifurcations. Asymptomatic carotids were associated with a lower 'skewness' and a lower 'kurtosis' than culprit and non-culprit carotid arteries.

With the restricted algorithms, similar trends were seen, primarily: (1) asymptomatic carotid arteries had a higher 'mean' than culprit carotid arteries, (2) asymptomatic carotid arteries had a higher 'entropy' than symptomatic carotid arteries and (3) asymptomatic carotid arteries had lower 'skewness' and 'kurtosis' than culprit carotid arteries. There were greater differences between asymptomatic versus culprit carotid arteries, than between asymptomatic and non-culprit carotid arteries, see Table 2.29.

Table 2.29 TexRAD Multi-slice analysis: Asx vs CC or NC

Texture Feature	Asymptomatic Artery – median (IQR)	Comparisons	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	238.8 (213.0, 289.2)	Asx vs CC	<0.0001*
		Asx vs NC	0.555
MPP	244.4 (218.8, 290.4)	Asx vs CC	<0.0001*
		Asx vs NC	0.5234
Standard Deviation	138.3 (114.7, 166.4)	Asx vs CC	0.194
		Asx vs NC	0.140
Entropy	5.85 (5.74, 5.94)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	-0.22 (-0.40, 0.07)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Kurtosis	-1.13 (-1.39, -0.61)	Asx vs CC	<0.0001*
		Asx vs NC	0.001*
Restricted I Algorithm [0, 200 HU]			
Mean	105.9 (101.3, 110.5)	Asx vs CC	<0.0001*
		Asx vs NC	0.3385

Standard Deviation	50.4 (47.7, 52.8)	Asx vs CC	<0.0001*
		Asx vs NC	0.003*
Entropy	5.14 (5.07, 5.17)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	0.10 (-0.01, 0.19)	Asx vs CC	<0.0001*
		Asx vs NC	0.081
Kurtosis	-0.97 (-1.03, -0.85)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Restricted II Algorithm [-1000, 200 HU]			
Mean	103.2 (95.2, 108.4)	Asx vs CC	0.006*
		Asx vs NC	0.732
MPP	106.1 (101.5, 110.6)	Asx vs CC	<0.0001*
		Asx vs NC	0.304
Standard Deviation	53.6 (48.9, 60.2)	Asx vs CC	<0.0001*
		Asx vs NC	0.051
Entropy	5.20 (5.11, 5.30)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	-0.07 (-0.23, 0.06)	Asx vs CC	<0.0001*
		Asx vs NC	0.188
Kurtosis	-0.67 (-0.79, -0.54)	Asx vs CC	0.002*
		Asx vs NC	0.111
Restricted III Algorithm [-1000, 300 HU]			
Mean	145.6 (138.5, 155.5)	Asx vs CC	<0.0001*
		Asx vs NC	0.367
MPP	149.1 (141.1, 161.9)	Asx vs CC	<0.0001*
		Asx vs NC	0.168
Standard Deviation	86.1 (77.6, 89.8)	Asx vs CC	<0.0001*
		Asx vs NC	0.022*
Entropy	5.59 (5.49, 5.64)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	0.13 (-0.07, 0.32)	Asx vs CC	<0.001*
		Asx vs NC	0.374
Kurtosis	-0.96 (-1.10, -0.86)	Asx vs CC	<0.0001*
		Asx vs NC	0.017*

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; MPP, mean of positive pixels; HU, Hounsfield units. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

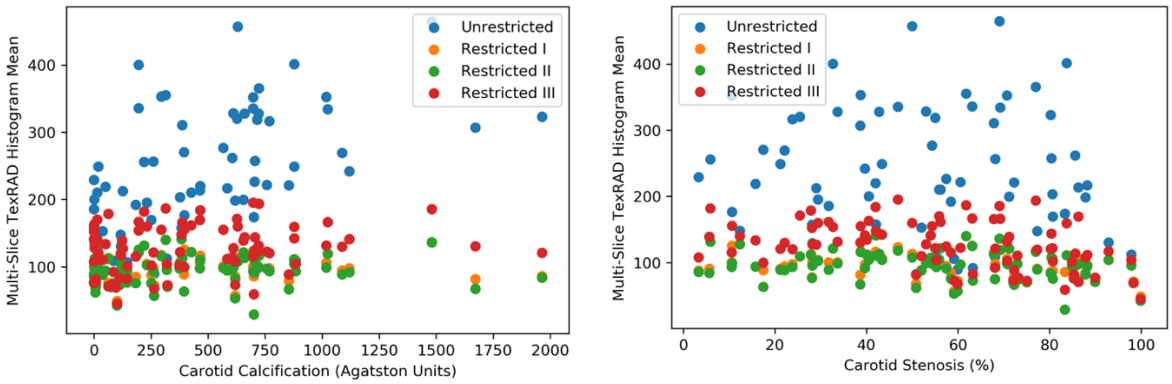


Figure 2.34 Scatter Plots of Multi-slice 'Mean' with Carotid Calcification and Stenosis

Table 2.30 Spearman's Rho correlation with Carotid Stenosis: Multi-slice approach

Texture Feature	Correlation	Spearman's rho and Significance
Unrestricted Algorithm [-1000, 1000 HU]		
Mean	With carotid stenosis	r_s = -0.318 p=0.004*
	With calcium	r_s = 0.611 p <0.0001*
Restricted I Algorithm [0, 200 HU]		
Mean	With carotid stenosis	r_s =-0.306 p =0.005*
	With calcium	r_s = 0.240 p =0.030*
Restricted II Algorithm [-1000, 200 HU]		
Mean	With carotid stenosis	r_s = -0.236 p =0.033*
	With calcium	r_s =0.182 p =0.102
Restricted III Algorithm [-1000, 300 HU]		
Mean	With carotid stenosis	r_s =-0.425 p <0.0001*
	With calcium	r_s =0.215 p =0.052

*r_s, Spearman's rho; *p-value<0.05.*

Texture features extracted in the multi-slice approach were not independent from carotid stenosis, see Figure 2.34. When no restriction was used (i.e. no resegmentation), there was a weak-to-moderate negative correlation between the texture feature 'mean' and the degree of carotid stenosis (r_s=-0.318, p=0.004), and there was a strong positive correlation with carotid calcium burden (r_s=0.611, p<0.0001). When resegmentation was applied with the restricted algorithms, these correlations were attenuated, see Table 2.30.

2.3.7.2.2 Comparisons with asymptomatic carotid plaque only

There were 28 asymptomatic carotid arteries with carotid atheroma visible on CT. Inter-patient comparisons between asymptomatic and symptomatic patients on a per-vessel basis when restricted to only these asymptomatic carotid arteries with carotid plaque revealed the same patterns as section 2.3.7.2.1. The consistent findings regardless of algorithm used were: (1) asymptomatic carotid arteries had a lower 'mean' and 'mpp' than culprit carotid arteries, (2) asymptomatic carotid arteries had a higher 'entropy' than culprit and non-culprit carotid arteries, (3) asymptomatic carotids were associated with a lower 'skewness' and a lower 'kurtosis' than symptomatic carotids, see Table 2.31.

Table 2.31 Multi-slice TexRAD features: Asx (plaque only) vs CC or NC

Texture Feature	Asymptomatic Artery – median (IQR)	Comparisons	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	246.4 (218.0, 301.3)	Asx vs CC	0.0002*
		Asx vs NC	0.289
Mean of Positive Pixels	250.4 (223.7, 304.2)	Asx vs CC	0.0004*
		Asx vs NC	0.276
Standard Deviation	150.1 (124.2, 178.4)	Asx vs CC	0.119
		Asx vs NC	0.795
Entropy	5.93 (5.79, 6.02)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	0.035 (-0.273, 0.255)	Asx vs CC	<0.0001*
		Asx vs NC	0.012*
Kurtosis	-0.730 (-1.15, -0.283)	Asx vs CC	<0.0001*
		Asx vs NC	0.114
Restricted I Algorithm [0, 200 HU]			
Mean	105.3 (101.2, 110.4)	Asx vs CC	0.002*
		Asx vs NC	0.465
Standard Deviation	50.8 (47.9, 52.8)	Asx vs CC	<0.0001*
		Asx vs NC	0.018*
Entropy	5.14 (5.07, 5.18)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	0.085 (0.008, 0.193)	Asx vs CC	<0.0001*
		Asx vs NC	0.153
Kurtosis	-0.990 (-1.03, -0.848)	Asx vs CC	<0.0001*
		Asx vs NC	0.001*
Restricted II Algorithm [-1000, 200 HU]			
Mean	100.3 (93.1, 109.2)	Asx vs CC	0.036*
		Asx vs NC	0.992
Mean of Positive Pixels	105.4 (101.3, 110.4)	Asx vs CC	0.002*
		Asx vs NC	0.465
Standard Deviation	54.1 (49.5, 60.9)	Asx vs CC	<0.0001*

		Asx vs NC	0.075
Entropy	5.21 (5.13, 5.31)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	-0.075 (-0.225, 0.018)	Asx vs CC	0.0004*
		Asx vs NC	0.180
Kurtosis	-0.645 (-0.783, -0.545)	Asx vs CC	0.019*
		Asx vs NC	0.384
Restricted III Algorithm [-1000, 300 HU]			
Mean	142.8 (136.7, 156.6)	Asx vs CC	<0.0001*
		Asx vs NC	0.803
Mean of Positive Pixels	148.5 (139.6, 161.0)	Asx vs CC	<0.0001*
		Asx vs NC	0.478
Standard Deviation	86.4 (80.2, 91.7)	Asx vs CC	<0.0001*
		Asx vs NC	0.038*
Entropy	5.61 (5.53, 5.65)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	0.125 (-0.018, 0.335)	Asx vs CC	<0.0001*
		Asx vs NC	0.549
Kurtosis	-0.895 (-1.11, -0.853)	Asx vs CC	<0.0001*
		Asx vs NC	0.067

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; HU, Hounsfield units. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

2.3.7.2.3 Carotid Plaque Type: Subset Analysis

For subset analysis according to carotid plaque type for inter-patient comparisons, there were too few noncalcified plaques (Asx: 1, CC: 2, NC: 5 carotid arteries) for stratified analysis of that plaque type. For subset analysis of calcified plaques, 14 asymptomatic carotid arteries, 9 culprit and 15 non-culprit carotid arteries were compared. For mixed plaque subset analysis, there were 13 asymptomatic carotid arteries, 30 culprit and 21 non-culprit carotid arteries.

For calcified plaques, using the unrestricted algorithm, the following trends were seen: (1) asymptomatic carotids had a lower 'standard deviation' than non-culprit carotids, (2) asymptomatic carotids had a higher 'entropy' than culprit and non-culprit carotid arteries, (3) asymptomatic carotids had a lower 'skewness' and a lower 'kurtosis' than symptomatic carotid arteries. However, following resegmentation these differences disappeared, except 'entropy' remained higher in asymptomatic carotid arteries compared with symptomatic carotid arteries.

For mixed plaques, using the unrestricted algorithm, the following trends were seen: (1) asymptomatic carotids had a higher 'mean' than culprit and non-culprit carotids, (2) asymptomatic carotids had a higher 'mpp' than culprit carotids, (3) asymptomatic carotids had a higher 'entropy' than symptomatic carotids and (4) asymptomatic carotids had a lower 'skewness' than culprit carotid arteries. These trends were maintained following resegmentation using the restricted algorithms (I-III), with the addition that asymptomatic carotids had a higher 'standard deviation' than culprit carotid arteries, see Table 2.32.

Table 2.32 Multi-slice Carotid Plaque Type Subset Analysis: Asx vs CC or NC

Texture Feature	**Plaque Type**	**Asymptomatic Artery**	**Culprit Artery**	**Non-culprit Artery**	**Comparisons**	***p*-value**
		Median (IQR)				
Unrestricted Algorithm [-1000, 1000 HU]						
Mean	Calcified	279.2 (223.9, 305.3)	334,4 (185.0, 365.4)	320.3 (263.3, 343.6)	Asx vs CC	0.395
					Asx vs NC	0.131
	Mixed	229.9 (218.3, 271.6)	121.5 (90.2, 202.6)	210.5 (176.5, 241.9)	Asx vs CC	0.0003*
					Asx vs NC	0.043*
MPP	Calcified	286.1 (226.9, 317.2)	334.4 (186.0, 366.3)	322.7 (278.6, 346.3)	Asx vs CC	0.472
					Asx vs NC	0.112

	Mixed	239.4 (219.4, 276.8)	122.4 (96.1, 215.9)	210.5 (185.0, 248.5)	Asx vs CC	0.0008*
					Asx vs NC	0.056
Standard Deviation	Calcified	170.4 (114.5, 187.2)	235.3 (137.6, 258.9)	203.4 (185.2, 218.2)	Asx vs CC	0.107
					Asx vs NC	0.008*
	Mixed	141.5 (131.0, 164.8)	100.7 (64.5, 184.8)	145.2 (102.0, 166.1)	Asx vs CC	0.055
					Asx vs NC	0.596
Entropy	Calcified	5.90 (5.69, 6.05)	5.07 (4.91, 5.26)	5.36 (5.31, 5.47)	Asx vs CC	<0.0001*
					Asx vs NC	0.0001*
	Mixed	5.94 (5.85, 6.02)	4.88 (4.76, 5.03)	5.24 (5.05 5.39)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
Skewness	Calcified	-0.055 (-0.340, 0.090)	0.900 (0.585, 1.22)	0.590 (0.345, 1.11)	Asx vs CC	0.0002*
					Asx vs NC	0.001*
	Mixed	0.140 (-0.140, 1.54)	1.32 (0.820, 1.63)	1.12 (-0.090, 1.68)	Asx vs CC	0.019*
					Asx vs NC	0.347
Kurtosis	Calcified	-1.12 (-1.27, -0.35)	0.040 (-0.780, 2.99)	0.220 (-0.715, 1.32)	Asx vs CC	0.011*
					Asx vs NC	0.010*
	Mixed	-0.470 (-0.810, 4.65)	1.54 (0.745, 3.97)	0.94 (-0.88, 5.65)	Asx vs CC	0.070
					Asx vs NC	0.395
Restricted I Algorithm [0, 200 HU]						
Mean	Calcified	107.4 (101.3, 110.3)	105.9 (98.4, 118.1)	106.9 (98.6, 119.5)	Asx vs CC	0.928
					Asx vs NC	0.984
	Mixed	104.7 (101.3, 110.3)	92.2 (75.7, 102.7)	101.4 (94.1, 114.2)	Asx vs CC	0.004*
					Asx vs NC	0.617
Standard Deviation	Calcified	48.9 (45.2, 52.6)	47.6 (39.0, 47.7)	47.6 (43.8, 49.2)	Asx vs CC	0.197
					Asx vs NC	0.246
	Mixed	51.0 (49.1, 52.6)	43.8 (38.7, 47.1)	48.7 (46.9, 51.4)	Asx vs CC	<0.0001*
					Asx vs NC	0.089
Entropy	Calcified	5.14 (5.03, 5.18)	4.17 (3.89, 4.24)	4.33 (4.21, 4.63)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
	Mixed	5.12 (5.09, 5.18)	4.48 (4.28, 4.61)	4.61 (4.46, 4.69)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
Skewness	Calcified	0.085 (0.023, 0.170)	0.090 (0.080, 0.510)	0.120 (-0.090, 0.220)	Asx vs CC	0.490
					Asx vs NC	0.968

	Mixed	0.080 (-0.030, 0.190)	0.500 (0.265, 0.698)	0.310 (-0.060, 0.460)	Asx vs CC	0.0001*
					Asx vs NC	0.174
Kurtosis	Calcified	-0.900 (-1.008, -0.833)	-1.00 (-1.09, -0.560)	-0.720 (-0.960, -0.480)	Asx vs CC	0.826
					Asx vs NC	0.097
	Mixed	-1 (-1.03, -0.950)	-0.395 (-0.71, 0.113)	-0.730 (-0.990, -0.490)	Asx vs CC	<0.0001*
					Asx vs NC	0.028*
Restricted II Algorithm [-1000, 200 HU]						
Mean	Calcified	100.3 (94.6, 109.8)	101.8 (97.5, 116.8)	104.4 (93.1, 117.0)	Asx vs CC	0.638
					Asx vs NC	0.646
	Mixed	102.8 (96.6, 109.0)	90.6 (70.5, 101.3)	97.5 (91.8, 110.8)	Asx vs CC	0.021*
					Asx vs NC	0.749
MPP	Calcified	107.6 (101.5, 110.3)	105.9 (98.4, 118.1)	106.9 (98.4, 119.5)	Asx vs CC	0.928
					Asx vs NC	0.984
	Mixed	104.8 (101.3, 110.3)	92.2 (75.7, 102.7)	101.4 (94.1, 114.2)	Asx vs CC	0.003*
					Asx vs NC	0.617
Standard Deviation	Calcified	52.6 (46.2, 58.6)	47.6 (39.0, 54.5)	49.1 (45.4, 55.7)	Asx vs CC	0.095
					Asx vs NC	0.447
	Mixed	54.0 (50.3, 60.8)	45.8 (39.5, 50.0)	52.3 (50.0, 54.2)	Asx vs CC	0.0002*
					Asx vs NC	0.258
Entropy	Calcified	5.21 (5.05, 5.31)	4.26 (3.89, 4.36)	4.37 (4.22, 4.65)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
	Mixed	5.19 (5.13, 5.30)	4.52 (4.36, 4.63)	4.69 (4.53, 4.75)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
Skewness	Calcified	-0.055 (-0.198, 0.048)	0.080 (-0.250, 0.190)	-0.060 (-0.295, 0.045)	Asx vs CC	0.430
					Asx vs NC	0.881
	Mixed	-0.120 (-0.240, -0.020)	0.310 (-0.050, 0.598)	0.010 (-0.210, 0.210)	Asx vs CC	0.003*
					Asx vs NC	0.162
Kurtosis	Calcified	-0.610 (-0.680, -0.273)	-0.850 (-1.02, -0.480)	-0.530 (-0.830, -0.295)	Asx vs CC	0.412
					Asx vs NC	0.944
	Mixed	-0.640 (-0.820, -0.550)	-0.190 (-0.565, 0.438)	-0.570 (-0.810, -0.240)	Asx vs CC	0.003*
					Asx vs NC	0.384
Restricted III Algorithm [-1000, 300 HU]						
Mean	Calcified				Asx vs CC	0.726

		147.7 (137.4, 150.5)	131.7 (121.0, 166.3)	142.5 (125.9, 163.1)	Asx vs NC	0.682
	Mixed	139.6 (136.2, 165.6)	104.4 (78.5, 121.5)	140.3 (122.5, 161.7)	Asx vs CC	<0.0001*
					Asx vs NC	0.478
MPP	Calcified	150.4 (141.2, 156.1)	138.4 (122.9, 166.3)	144.9 (129.7, 163.8)	Asx vs CC	0.589
					Asx vs NC	0.529
	Mixed	147.7 (140.7, 167.4)	108.4 (83.2, 121.7)	144.8 (125.8, 167.3)	Asx vs CC	<0.0001*
					Asx vs NC	0.358
Standard Deviation	Calcified	86.1 (73.0, 90.5)	75.9 (66.4, 79.3)	77.4 (68.7, 85.0)	Asx vs CC	0.035*
					Asx vs NC	0.215
	Mixed	86.7 (82.8, 87.7)	64.8 (56.0, 72.9)	79.7 (75.2, 89.5)	Asx vs CC	<0.0001*
					Asx vs NC	0.066
Entropy	Calcified	5.61 (5.42, 5.65)	4.56 (4.35, 4.60)	4.76 (4.56, 5.08)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
	Mixed	5.60 (5.54, 5.65)	4.73 (4.46, 4.86)	4.97 (4.83, 5.03)	Asx vs CC	<0.0001*
					Asx vs NC	<0.0001*
Skewness	Calcified	0.230 (-0.005, 0.380)	0.170 (0.070, 0.840)	0.210 (0.050, 0.405)	Asx vs CC	0.549
					Asx vs NC	0.757
	Mixed	0.010 (-0.040, 0.150)	0.760 (0.390, 1.13)	0.090 (-0.210, 0.570)	Asx vs CC	0.0001*
					Asx vs NC	0.697
Kurtosis	Calcified	-0.875 (-1.06, -0.733)	-0.950 (-1.09, -0.030)	-0.750 (-0.845, -0.685)	Asx vs CC	0.873
					Asx vs NC	0.230
	Mixed	-1.01 (-1.16, -0.870)	0.320 (-0.320, 1.11)	-0.680 (-1.07, -0.420)	Asx vs CC	<0.0001*
					Asx vs NC	0.062

*MPP, mean of positive pixels; HU, Hounsfield units; Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries. For calcified plaque subset analysis, 14 asymptomatic carotid arteries were compared with 9 culprit carotid arteries and 15 non-culprit carotid arteries respectively. For mixed plaque subset analysis, 13 asymptomatic carotid arteries were compared with 30 culprit carotid arteries and 21 non-culprit carotid arteries respectively. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

2.3.7.2.4 Carotid Stenosis Severity: Subset Analysis

The number of carotid arteries with carotid stenosis between 1 up to 50%, there were 25 asymptomatic, 3 culprit and 27 non-culprit carotid arteries. In the above 50% carotid stenosis group, there were 3 asymptomatic, 38 culprit and 14 non-culprit carotid arteries. Since the sample size for culprit carotid arteries with carotid stenosis below 50% is small and for asymptomatic carotid arteries with stenosis above 50% is small, the interpretation of these comparisons is limited.

The comparisons were as follows:

1. Carotid stenosis: 1-50%
    a. Asx vs CC: 25 vs 3 carotid arteries
    b. Asx vs NC: 25 vs 27 carotid arteries
2. Carotid stenosis: ≥ 50%
    a. Asx vs CC: 3 vs 38 carotid arteries
    b. Asx vs NC: 3 vs 14 carotid arteries

The differences between carotid artery types in the below 50% and the above 50% carotid stenosis groups are largely similar. In both cases, using the unrestricted algorithm: (1) asymptomatic carotid arteries have a higher 'mean' and 'mpp' than culprit carotids and (2) asymptomatic carotid arteries have a higher 'entropy' than culprit and non-culprit carotid arteries. Similarly, following resegmentation with the restricted algorithms, asymptomatic carotid arteries consistently have a higher 'entropy' than culprit and non-culprit carotid arteries and asymptomatic carotid arteries frequently have a higher standard deviation than culprit carotid arteries, see Table 2.33.

Table 2.33 Multi-slice Carotid Stenosis Subset Analysis: Asx vs CC or NC

Texture Feature	**Carotid Stenosis**	**Asymptomatic Artery**	**Culprit Artery**	**Non-culprit Artery**	**Comparisons**	***p*-value**
		Median (IQR)				
Unrestricted Algorithm [-1000, 1000 HU]						
Mean	1-50%	229.9 (217.2, 297.0)	195.3 (165.7, 203.8)	248.8 (182.4, 318.4)	Asx vs CC	0.014*
					Asx vs NC	0.787
					Asx vs CC	0.044*

	≥ 50%	337.7 (302.4, 350.6)	121.5 (84.3, 249.6)	217.6 (204.3, 302.4)	Asx vs NC	0.068
MPP	1-50%	240.4 (219.4, 297.6)	196.9 (166.5, 205.1)	250.4 (197.0, 322.4)	Asx vs CC	0.004*
					Asx vs NC	0.772
	≥ 50%	348.6 (314.3, 359.6)	122.4 (90.6, 256.0)	221.3 (206.6, 303.9)	Asx vs CC	0.038*
					Asx vs NC	0.068
Standard Deviation	1-50%	144.8 (115.3, 169.8)	126.6 (95.9, 147.9)	155.5 (106.3, 195.6)	Asx vs CC	0.391
					Asx vs NC	0.535
	≥ 50%	249.5 (219.7, 251.8)	105.0 (55.1, 223.6)	165.9 (140.7, 205.3)	Asx vs CC	0.086
					Asx vs NC	0.047*
Entropy	1-50%	5.91 (5.75, 5.94)	5.02 (4.73, 5.10)	5.32 (5.07, 5.44)	Asx vs CC	0.001*
					Asx vs NC	<0.0001*
	≥ 50%	6.47 (6.41, 6.49)	4.90 (4.62, 5.07)	5.26 (5.17, 5.40)	Asx vs CC	0.0002*
					Asx vs NC	0.003*
Skewness	1-50%	-0.06 (-0.28, 0.11)	1.49 (1.19, 1.53)	0.47 (-0.065, 1.21)	Asx vs CC	0.032*
					Asx vs NC	0.019*
	≥ 50%	0.66 (0.60, 0.67)	1.19 (0.658, 1.58)	1.07 (0.245, 1.24)	Asx vs CC	0.147
					Asx vs NC	0.509
Kurtosis	1-50%	-0.93 (-1.16, -0.33)	2.99 (1.45, 3.82)	-0.19 (-1.26, 2.05)	Asx vs CC	0.062
					Asx vs NC	0.204
	≥ 50%	-0.23 (-0.35, 0.12)	1.29 (0.163, 3.12)	0.885 (-0.71, 1.40)	Asx vs CC	0.147
					Asx vs NC	0.509
Restricted I Algorithm [0, 200 HU]						
Mean	1-50%	105.8 (101.2, 110.7)	104.2 (103.4, 108.3)	101.4 (94.4, 117.1)	Asx vs CC	0.780
					Asx vs NC	0.726
	≥ 50%	103.1 (102.0, 105.6)	92.2 (75.7, 105.5)	102.9 (97.6, 107.1)	Asx vs CC	0.178
					Asx vs NC	0.768
Standard Deviation	1-50%	49.3 (47.3, 52.6)	43.1 (41.8, 43.2)	48.3 (46.0, 50.3)	Asx vs CC	0.007*
					Asx vs NC	0.153
	≥ 50%	54.1 (53.0, 55.2)	44.4 (38.2, 47.6)	47.7 (44.7, 49.0)	Asx vs CC	0.004*
					Asx vs NC	0.021*
Entropy	1-50%	5.15 (5.09, 5.18)	4.38 (4.28, 4.42)	4.61 (4.33, 4.69)	Asx vs CC	0.001*
					Asx vs NC	<0.0001*

	≥ 50%	5.07 (5.05, 5.13)	4.41 (4.17, 4.60)	4.50 (4.40, 4.60)	Asx vs CC	0.0002*
					Asx vs NC	0.003*
Skewness	1-50%	0.09 (0.01, 0.19)	0.33 (0.30, 0.43)	0.12 (-0.085, 0.33)	Asx vs CC	0.002*
					Asx vs NC	0.562
	≥ 50%	0.00 (-0.07, 0.11)	0.465 (0.135, 0.678)	0.33 (0.07, 0.45)	Asx vs CC	0.067
					Asx vs NC	0.197
Kurtosis	1-50%	-0.95 (-1.03, -0.84)	-0.56 (-0.64, -0.56)	-0.72 (-0.915, -0.50)	Asx vs CC	0.004*
					Asx vs NC	0.004*
	≥ 50%	-1.03 (-1.09, -1.03)	-0.395 (-0.863, 0.235)	-0.67 (-1.00, -0.43)	Asx vs CC	0.190*
					Asx vs NC	0.091
Restricted II Algorithm [-1000, 200 HU]						
Mean	1-50%	102.8 (94.4, 109.6)	102.4 (102.1, 107.4)	98.7 (89.3, 112.5)	Asx vs CC	0.572
					Asx vs NC	0.968
	≥ 50%	95.0 (92.0, 96.3)	90.58 (71.1, 102.9)	98.3 (95.0, 106.7)	Asx vs CC	0.652
					Asx vs NC	0.362
MPP	1-50%	106.0 (101.3, 110.7)	104.2 (103.4, 108.3)	101.4 (94.4, 117.1)	Asx vs CC	0.738
					Asx vs NC	0.726
	≥ 50%	103.4 (101.9, 105.8)	92.2 (75.7, 105.5)	102.9 (97.4, 107.1)	Asx vs CC	1.78
					Asx vs NC	0.768
Standard Deviation	1-50%	52.7 (48.4, 58.9)	43.9 (42.3, 45.1)	51.7 (48.5, 56.3)	Asx vs CC	0.019*
					Asx vs NC	0.430
	≥ 50%	65.7 (61.5, 66.7)	45.8 (38.5, 51.6)	50.2 (46.0, 53.7)	Asx vs CC	0.002*
					Asx vs NC	0.012*
Entropy	1-50%	5.22 (5.12, 5.30)	4.39 (4.28, 4.44)	4.64 (4.40, 4.75)	Asx vs CC	0.001*
					Asx vs NC	<0.0001*
	≥ 50%	5.19 (5.18, 5.28)	4.47 (4.27, 4.61)	4.55 (4.41, 4.67)	Asx vs CC	0.0002*
					Asx vs NC	0.003*
Skewness	1-50%	-0.03 (-0.21, 0.04)	0.19 (0.05, 0.355)	-0.06 (-0.3, 0.12)	Asx vs CC	0.125
					Asx vs NC	0.779
	≥ 50%	-0.24 (-0.46, -0.18)	0.20 (-0.068, 0.518)	0.025 (-0.08, -.278)	Asx vs CC	0.038*
					Asx vs NC	0.032*
Kurtosis	1-50%	-0.66 (-0.82, -0.55)	-0.48 (-0.595, -0.005)	-0.49 (-0.85, -0.15)	Asx vs CC	0.248
					Asx vs NC	0.280

	≥ 50%	-0.59 (-0.61, -0.02)	-0.23 (-0.805, 0.323)	-0.53 (-0.75, -0.28)	Asx vs CC	0.795
					Asx vs NC	0.591
Restricted III Algorithm [-1000, 300 HU]						
Mean	1-50%	149.4 (138.5, 159.5)	134.6 (131.0, 147.3)	142.5 (130.3, 158.4)	Asx vs CC	0.524
					Asx vs NC	0.826
	≥ 50%	136.2 (134.6, 136.6)	104.4 (77.3, 124.0)	134.5 (121.5, 164.8)	Asx vs CC	0.108
					Asx vs NC	1.000
MPP	1-50%	151.5 (139.6, 163.5)	134.6 (131.5, 148.1)	145.8 (133.5, 168.4)	Asx vs CC	0.219
					Asx vs NC	0.660
	≥ 50%	146.7 (142.9, 148.0)	108.4 (82.5, 126.8)	137.1 (124.1, 165.9)	Asx vs CC	0.067
					Asx vs NC	0.859
Standard Deviation	1-50%	86.2 (77.6, 87.9)	71.7 (67.6, 79.0)	78.7 (71.6, 87.8)	Asx vs CC	0.106
					Asx vs NC	0.165
	≥ 50%	91.4 (87.9, 92.6)	66.2 (53.6, 75.5)	78.5 (68.7, 88.0)	Asx vs CC	0.003*
					Asx vs NC	0.068
Entropy	1-50%	5.61 (5.53, 5.65)	4.63 (4.52, 4.80)	4.91 (4.72, 5.08)	Asx vs CC	0.001*
					Asx vs NC	<0.0001*
	≥ 50%	5.60 (5.57, 5.66)	4.57 (4.40, 4.78)	4.85 (4.76, 5.03)	Asx vs CC	0.0002*
					Asx vs NC	0.003*
Skewness	1-50%	0.13 (-0.01, 0.35)	0.84 (0.49, 0.85)	0.15 (-0.155, 0.405)	Asx vs CC	0.062
					Asx vs NC	0.928
	≥ 50%	0.01 (-0.09, 0.10)	0.695 (0.180, 1.08)	0.275 (0.02, 0.593)	Asx vs CC	0.051
					Asx vs NC	0.300
Kurtosis	1-50%	-0.90 (-1.11, -0.86)	-0.15 (-0.675, -0.09)	-0.77 (-1.04, -0.665)	Asx vs CC	0.248
					Asx vs NC	0.036*
	≥ 50%	-0.81 (-0.94, -0.64)	0.215 (-0.67, 1.2)	-0.775 (-1.16, -0.343)	Asx vs CC	0.086
					Asx vs NC	0.953

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries; HU, Hounsfield units; MPP, mean of positive pixels. In the category 1-50%, there were 25 asymptomatic, 3 culprit and 27 non-culprit carotid arteries. In the category ≥ 50%, there were 3 asymptomatic, 38 culprit and 14 non-culprit carotid arteries. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

2.3.8 Unenhanced CT

2.3.8.1 Level of agreement between CTA and unenhanced CT TexRAD features

There was poor agreement in absolute terms and in consistency between texture features extracted from CT angiography scans and unenhanced CT scans, as quantified by the intraclass correlation coefficient in Table 2.34. Only the texture feature 'standard deviation' had moderate agreement between the two scan types, with an ICC >0.5. Using resegmentation with the restricted algorithm and limiting the range of Hounsfield units used to calculate the texture features did not improve the level of agreement between CTA and unenhanced CT.

Table 2.34 Intraclass Correlation Coefficients between TexRAD texture features in CT angiography and Unenhanced CT

Unrestricted Algorithm	**ICC absolute agreement**	**ICC consistency**
Mean	0.154 (-0.083, 0.407)	0.385 (0.184, 0.555)
MPP	0.171 (-0.086, 0.438)	0.415 (0.219, 0.579)
Standard deviation	0.578 (0.022, 0.806)	0.739 (0.622, 0.824)
Entropy	0.313 (0.102, 0.496)	0.310 (0.101, 0.493)
Skewness	0.102 (-0.063, 0.280)	0.186 (-0.031, 0.386)
Kurtosis	0.062 (-0.074, 0.217)	0.106 (-0.113, 0.314)
Restricted I Algorithm	**ICC absolute agreement**	**ICC consistency**
Mean	0.041 (-0.044, 0.154)	0.175 (-0.043, 0.376)
MPP	0.041 (-0.044, 0.154)	0.175 (-0.043, 0.376)
Standard deviation	0.121 (-0.070, 0.322)	0.253 (0.040, 0.445)
Entropy	-0.147 (-0.329, 0.060)	-0.220 (-0.416, -0.005)
Skewness	0.004 (-0.085, 0.117)	0.010 (-0.207, 0.225)
Kurtosis	0.110 (-0.067, 0.298)	0.222 (0.006, 0.417)

MPP, mean of positive pixels; ICC, intraclass correlation coefficients. Single measures ICC and 95% confidence interval reported; ICCs calculated using culprit and non-culprit carotid multi-slice texture features

2.3.8.2 Unenhanced CT correlation with carotid calcium

In unenhanced CT, when using the unrestricted algorithm, the texture feature 'mean' was strongly positively correlated with carotid calcium burden (r_s=0.897, p<0.0001). With the restricted I algorithm [0, 200 HU], the texture feature remained strongly positively correlated with carotid calcium (r_s=0.748, p<0.0001). With the restricted II algorithm [-1000, 200 HU], there remained a moderate positive correlation with calcium (r_s=0.557, p<0.0001) and with the restricted III algorithm [-1000, 300 HU], there remained a strong positive correlation with calcium as well (r_s=0.711, p<0.0001). Therefore, resegmentation as determined by the restricted algorithms developed for CTA scans, was not directly applicable to unenhanced CT scans.

2.3.8.3 Symptomatic Patients: Culprit versus Non-culprit Carotids

In the comparative analysis of TexRAD texture features from unenhanced CT, there were no differences between culprit and non-culprit carotid arteries, regardless of the algorithm used, see Table 2.35.

Table 2.35 Multi-slice Unenhanced CT Features: Culprit versus Non-culprit

Texture Features	Culprit Artery – median (IQR)	Non-culprit Artery – median (IQR)	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	65.4 (54.9, 96.5)	78.2 (55.8, 94.8)	0.758
Mean of Positive Pixels	67.3 (58.4, 103.0)	84.7 (59.6, 106.8)	0.555
Standard Deviation	75.5 (47.3, 142.9)	95.8 (51.8, 140.5)	0.564
Entropy	5.13 (4.76, 5.38)	5.11 (4.88, 5.40)	0.216
Skewness	3.11 (2.20, 3.99)	2.99 (2.47, 3.74)	0.892
Kurtosis	13.1 (6.01, 25.5)	11.8 (7.09, 16.5)	0.929
Restricted I Algorithm [0, 200 HU]			
Mean	59.1 (53.9, 66.0)	60.2 (55.4, 64.8)	0.595
Standard Deviation	34.5 (26.7, 39.3)	34.0 (27.3, 38.2)	0.106
Entropy	4.77 (4.51, 4.85)	4.74 (4.53, 4.87)	0.315
Skewness	1.05 (0.79, 1.33)	1.08 (0.61, 1.30)	0.216
Kurtosis	1.55 (0.90, 3.07)	1.45 (0.50, 2.44)	0.249
Restricted II Algorithm [-1000, 200 HU]			
Mean	54.9 (50.3, 62.2)	54.8 (49.3, 62.0)	0.816
Mean of Positive Pixels	59.3 (54.0, 66.3)	60.2 (56.4, 65.7)	0.627
Standard Deviation	39.2 (31.3, 44.5)	39.4 (32.4, 43.9)	0.158
Entropy	4.92 (4.64, 5.04)	4.88 (4.71, 5.04)	0.160
Skewness	0.52 (0.07, 0.80)	0.28 (-0.13, 0.75)	0.118

Kurtosis	1.39 (0.86, 2.29)	1.46 (1.16, 2.56)	0.599
Restricted III Algorithm [-1000, 300 HU]			
Mean	58.8 (50.4, 71.4)	58.3 (51.6, 68.9)	0.808
Mean of Positive Pixels	62.8 (55.2, 74.9)	65.3 (57.0, 72.3)	0.521
Standard Deviation	48.0 (36.3, 59.9)	47.8 (40.5, 56.6)	0.193
Entropy	5.02 (4.69, 5.16)	4.97 (4.78, 5.19)	0.168
Skewness	1.44 (0.60, 1.73)	1.37 (0.69, 1.72)	0.389
Kurtosis	3.58 (1.72, 6.41)	3.62 (2.63, 5.08)	0.954

SD, standard deviation; IQR, interquartile range. The P value was obtained using the Wilcoxon signed-rank test, after assessing for normality of differences or a paired t-test as appropriate.

2.3.8.4 Inter-patient comparisons: Asymptomatic versus Symptomatic

There were differences however, between asymptomatic carotid arteries versus culprit carotids and asymptomatic versus non-culprit carotid arteries respectively, see Table 2.36.

Table 2.36 TexRAD Multi-slice analysis in Unenhanced CT: Asx vs CC or NC

Texture Feature	Asymptomatic Artery – median (IQR)	Comparisons	*p*-value
Unrestricted Algorithm [-1000, 1000 HU]			
Mean	48.7 (42.9, 59.5)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Mean of Positive Pixels	52.6 (47.7, 61.7)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Standard Deviation	30.8 (22.0, 53.7)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Entropy	4.45 (4.30, 4.80)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	-0.35 (-1.22, 1.02)	Asx vs CC	<0.0001*
		Asx vs NC	0.003*
Kurtosis	3.50 (1.99, 22.6)	Asx vs CC	0.028*
		Asx vs NC	0.026*
Restricted I Algorithm [0, 200 HU]			
Mean	50.8 (47.5, 55.1)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Standard Deviation	19.7 (17.5, 27.5)	Asx vs CC	<0.001*
		Asx vs NC	<0.0001*
Entropy	4.25 (4.17, 4.35)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	0.40 (-0.25, 1.23)	Asx vs CC	<0.0001*
		Asx vs NC	0.003*
Kurtosis	1.18 (0.04, 3.43)	Asx vs CC	0.402
		Asx vs NC	0.513
Restricted II Algorithm [-1000, 200 HU]			
Mean	47.9 (42.2, 52.7)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*

Mean of Positive Pixels	50.8 (47.6, 55.3)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Standard Deviation	26.2 (22.0, 33.9)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Entropy	4.45 (4.29, 4.73)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	-0.55 (-1.23, 0.32)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Kurtosis	2.74 (1.76, 3.83)	Asx vs CC	0.001*
		Asx vs NC	0.003*
Restricted III Algorithm [-1000, 300 HU]			
Mean	48.3 (43.0, 55.7)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Mean of Positive Pixels	52.6 (47.7, 58.9)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Standard Deviation	29.1 (22.1, 41.2)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Entropy	4.46 (4.30, 4.77)	Asx vs CC	<0.0001*
		Asx vs NC	<0.0001*
Skewness	-0.29 (-1.22, 1.37)	Asx vs CC	<0.0001*
		Asx vs NC	0.001*
Kurtosis	3.52 (2.01, 6.38)	Asx vs CC	0.771
		Asx vs NC	0.968

*Asx, asymptomatic carotid arteries; CC, culprit carotid arteries; NC, non-culprit carotid arteries. Comparisons: Asx vs CC or Asx vs NC: Mann-Whitney U-test. *p-value<0.05.*

2.4 Discussion

2.4.1 Summary of key findings

2.4.1.1 Summary of key findings: Carotid Characteristics

There was no difference in the degree of carotid calcification between culprit versus non-culprit carotid arteries, either on a slice-by-slice basis or on a total calcium score per carotid artery basis. The spatial distribution of the carotid calcium score in the carotid artery was concentrated around the carotid bifurcation in both culprit and non-culprit arteries. However, when comparing symptomatic and asymptomatic patients, the latter group had significantly lower levels of carotid calcification. Nevertheless, the calcium distribution was similar to that found in symptomatic patients, centred around the carotid bifurcation.

The mean degree carotid luminal stenosis was higher in culprit carotid arteries compared with non-culprit and asymptomatic carotid arteries, respectively. This likely reflects the inclusion criteria of the vascular imaging studies from which this carotid dataset was drawn. For example, the inclusion criteria for the CHAI study mandated a greater than 30% stenosis in the *culprit* carotid artery and the inclusion criteria for the ICARUSS study required at least a 50% stenosis in the *culprit* carotid artery.

2.4.1.2 Summary of key findings: TexRAD Texture Analysis

The chapter establishes that there are important differences in the CT angiography TexRAD texture profiles of asymptomatic, culprit and non-culprit carotid arteries in both single-slice and multi-slice analyses. The differences between asymptomatic and culprit carotid arteries are greater than those between asymptomatic and non-culprit carotid arteries. Resegmentation, the process of restricting the range of Hounsfield units used to derive texture features within the ROI i.e. by using a 'restricted' algorithm, was necessary in certain instances to reveal these differences. The most important findings in TexRAD texture analysis have been summarised in Table 2.37.

Table 2.37 Relative CTA texture feature differences between carotid artery types

Unrestricted Algorithm	Single-slice analysis			Multi-slice analysis		
	Asx	CC	NC	Asx	CC	NC
Mean	↑↑	-	-	↑↑	-	↑↑
MPP	↑↑	-	-	↑↑	-	↑↑
SD	-	-	-	-	-	-
Entropy	↓	-	-	↑↑	-	↑
Skewness	↓↓	-	-	↓↓	-	↓
Kurtosis	↓↓	-	-	↓↓	-	↓
Restricted Algorithms	**Single-slice analysis**			**Multi-slice analysis**		
	Asx	**CC**	**NC**	**Asx**	**CC**	**NC**
Mean	↑↑	-	↑	↑↑	-	↑↑
MPP	↑↑	-	↑↑	↑↑	-	↑↑
SD	↑	-	↑	↑↑	-	↑
Entropy	↓	-	-	↑↑	-	↑
Skewness	↓↓	-	↓↓	↓↓	-	↓↓
Kurtosis	↓	-	↓	↓↓	-	↓

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries; MPP, mean of positive pixels; SD, standard deviation. The table summarises the relative differences in TexRAD (first order) texture features (listed in the left-hand column) between asymptomatic, culprit and non-culprit carotid arteries. The findings are summarised for single-slice and multi-slice analysis, as well as for the unrestricted algorithm (no resegmentation applied) and for the restricted algorithms (resegmentation applied). The colours and symbols indicate the relative differences in magnitude of the texture features between the carotid arteries: -, no difference; ↑↑, higher; ↑, slightly higher; ↓↓, lower; ↓, slightly lower.

2.4.2 Interpretation of findings

2.4.2.1 CT angiography: Single-slice approach

When using a single-slice approach (comparing an axial slice at the carotid bifurcation) in the comparative analysis of culprit versus non-culprit carotid arteries, resegmentation was necessary to reveal differences in TexRAD first order texture features. The three restricted algorithms investigated in this study had an upper limit of 200 or 300 HU, this limited the effect of calcification and luminal contrast on the extracted texture features. When using the unrestricted algorithm, it is likely the high grey level values from carotid calcification dominated the image histogram, such that differences in carotid plaque and tissue composition may be obscured. Subsequently, since there were no differences in calcium burden between culprit and non-culprit carotid arteries, we could not detect differences between culprit and non-culprit carotid arteries using first order texture features.

The most consistent differences in TexRAD texture features between culprit versus non-culprit carotid arteries when using the restricted algorithms (I-III) were: (1) culprit carotid arteries had a lower 'mean of positive pixels' than non-culprit carotid arteries and, related to this, (2) culprit carotid arteries have a more positive 'skewness'. Possible explanations for this difference could be the greater extent of lipid and intraplaque haemorrhage, both associated with low Hounsfield units, known to be present in culprit plaques. We excluded the degree of carotid stenosis as a confounder.

There were robust differences in first order texture features between asymptomatic carotid arteries and culprit or non-culprit carotid arteries respectively, regardless of algorithm used and accounting for differences in plaque burden. In the unrestricted algorithm, this likely reflects the lower calcium burden associated with asymptomatic carotid arteries. In the restricted algorithms, fewer differences were observed compared with the unrestricted algorithm as the effect of carotid calcium was controlled for. The first order texture feature profile of asymptomatic carotid arteries and non-culprit carotid arteries were similar, except that non-culprit carotid arteries had a higher 'entropy' in single-slice analysis. Interestingly, this association with 'entropy' was reversed in multi-slice analysis (discussed further below).

When using the LoG filter of size 2 mm for image pre-processing prior to texture feature extraction, there were no differences culprit versus non-culprit carotid arteries, regardless of the algorithm used. This likely relates to the difference in size between carotid arteries and carotid plaque versus tumours for which these filters were originally designed. Since unfiltered features already demonstrated statistical differences between carotid artery types, we focused on unfiltered features only within this thesis. An additional benefit of using unfiltered texture features is that the first order texture features can be interpreted in relation to Hounsfield units, whereas filtered features are no longer calibrated to this scale. However, future work could investigate smaller sigma sizes for the LoG filter, for example 0.25, 0.5, 0.75, and 1 mm. For example, Dolotova et al. in their application of radiomics to CTA scans assessing the 'vesselness' of intracranial blood vessels, applied a Gaussian filter prior to texture feature extraction to reduce image noise and used a sigma filter of size 0.7 mm according to the resolution of their CT scanner[96].

2.4.2.2 CT angiography Multi-slice approach

The single-slice approach above, in which the carotid bifurcation was used as a consistent anatomical landmark for ROI placement, was methodologically straightforward and highly reproducible. However, carotid atherosclerosis is not confined to a single slice. Therefore, a multi-slice radiomic analysis, derived from a 3D volume of carotid artery was undertaken to test whether there were important differences missed when using a single slice approach.

Consistent with our original hypothesis, the multi-slice approach was beneficial in revealing differences in first order texture features between all carotid artery types compared with the single-slice approach. This agrees with the literature, for example, Ng et al. 2013[170] found that whole tumour CT analysis was more representative of tumour heterogeneity than analysis of only the single largest cross-sectional area.

However, whilst the texture features derived in single-slice analysis were independent of carotid stenosis and extent of carotid calcification following resegmentation (using the restricted algorithms), there remained a weak-to-moderate correlation between texture features derived from multi-slice analysis and extent of carotid stenosis and calcification.

Furthermore, a surprising consistent finding in multi-slice analysis was the higher 'entropy' noted in asymptomatic carotid arteries versus non-culprit carotid arteries and culprit carotid arteries (whether asymptomatic carotid arteries with plaque only were considered or all asymptomatic carotid arteries). TexRAD 'entropy' is considered a quantitative measure of heterogeneity[171,172] and has been found to be increased in clear cell renal cancers versus papillary types, as well as high-grade renal cancers[172]. This suggests that within the carotid artery volume, the components of the asymptomatic carotid artery (not limited to the presence of visible plaque on CT) are more heterogeneous compared with symptomatic carotid arteries, rejecting our original hypothesis that culprit carotid arteries would display the most heterogeneity. The multi-slice findings suggest a gradation in heterogeneity with asymptomatic carotid arteries having the highest 'entropy', followed by non-culprit carotid arteries and then culprit carotid arteries having the lowest 'entropy'.

2.4.2.3 CT angiography Subset Analysis

When stratifying by carotid plaque type in the intra-patient comparison of culprit versus non-culprit carotid arteries, there were no longer any differences between them for paired calcified carotid plaques. This is not surprising since there were no differences in carotid calcium score nor carotid calcium spatial distributions between culprit and non-culprit carotid arteries. However, the differences remained for carotid arteries with mixed plaque, suggesting that texture features are influenced by (and reflective of) carotid atherosclerotic plaque composition. The sample size for noncalcified carotid plaque (2 pairs) was too small to conclusively identify any differences.

For inter-patient comparisons of asymptomatic versus symptomatic patients, carotid plaque subset analysis was only possible for calcified and mixed plaque types as there was only 1 asymptomatic carotid artery with noncalcified plaque. In the multi-slice approach, both asymptomatic calcified and mixed plaque types displayed higher 'entropy' than culprit and non-culprit carotid arteries in all algorithms (unrestricted and restricted I-III). It could be argued that this difference could be due to lower carotid stenosis in asymptomatic carotid arteries compared with culprit and non-culprit carotid arteries. However, in the subset analysis according to carotid stenosis category (below 50% or above 50%) to control for differences in carotid stenosis between carotid artery types, asymptomatic carotid arteries consistently had a higher 'entropy' than culprit and non-culprit carotid arteries regardless of algorithm used. Therefore, these texture features are capturing information regarding carotid artery status beyond luminal stenosis.

2.4.2.4 Unenhanced CT versus CT angiography

Using unenhanced CT-derived texture features, there were statistically significant differences between the carotid arteries of asymptomatic patients versus symptomatic patients. However, there were no differences between unenhanced CT carotid radiomic features of culprit versus non-culprit carotid arteries in symptomatic patients. Texture features derived from unenhanced CT were strongly positively correlated with carotid calcium burden, even after resegmentation (use of restricted algorithms). Therefore, differences in first order texture features derived from unenhanced CT likely reflected differences in carotid calcium burden between asymptomatic carotid arteries and those from symptomatic patients.

The level of agreement between TexRAD texture features extracted from CT angiography and unenhanced CT was assessed using the intraclass correlation coefficient for absolute agreement and consistency (single measures reported). There was poor agreement for all texture features (ICC <0.5), except 'standard deviation' which demonstrated moderate agreement. This demonstrates that texture features are specific to imaging modality and the administration of contrast. This is reflected by the ineffectiveness of the restricted algorithms in improving the detection of differences in texture features from unenhanced CT, versus CT angiography scans as the restricted algorithms were designed with CT angiography in mind.

2.4.3 Limitations

The texture features derived from TexRAD are first order features. This involves extracting the histogram of the grey level values within an ROI and calculating statistics based on them, for example, the 'mean' of the histogram or the 'standard deviation' of the histogram. Consequently, first order texture features do not consider the spatial relationships between pixels, missing an opportunity to capture potentially important information. We address this limitation by investigating higher-order texture features using PyRadiomics in the next chapter.

Another limitation is the manual segmentation of the ROIs in this study; a laborious process susceptible to intra-observer and inter-observer variability. In addition, since the ROI encompassed the whole carotid artery and not just the carotid plaque, texture features may not be reflecting characteristics of only the plaque, but all the components of the carotid artery within that ROI that contribute to the image histogram. In Chapter 4, we will investigate the effects of intra-observer and inter-observer variability on radiomic features to identify a set of robust radiomic features.

In addition, 8 CTA scans received an image quality score of 0 and were excluded from texture analysis. Scans with an image quality score of 0 were deemed to be of insufficient quality for radiomic analysis due to excessive calcification or other imaging artefacts. Although we had investigated the restricted algorithms which can limit the impact of carotid calcification on the texture features, scans of image quality 0 were obscured to the extent it was difficult to manually delineate the ROIs around the carotid artery accurately. ROI segmentation is a step that takes place prior to texture feature calculation and so resegmentation with the restricted algorithms would not have helped in this case.

2.5 Conclusions

In this chapter, we conclude that texture analysis is feasible for carotid CTA studies. There are several differences between culprit and non-culprit carotid arteries within symptomatic patients, in both single-slice and multi-slice analysis, unmasked when carotid calcium is aggressively excluded with resegmentation (i.e. using a restricted algorithm).

In comparison, when comparing carotid arteries from asymptomatic patients with those from symptomatic patients, there are also clear differences in texture features, present without the need for resegmentation. Multi-slice analysis is beneficial in detecting texture feature differences between carotid artery types, however, first order TexRAD features are weakly correlated with carotid stenosis and carotid calcification, whilst single-slice analysis features are independent.

Chapter 3: Radiomic analysis of Carotid CT Imaging with PyRadiomics

Chapter summary:

The chapter provides details of the methods, results and a discussion in relation to the statistical differences in radiomic features (first order and higher-order radiomic features) extracted using PyRadiomics in single-slice and multi-slice analysis of CT scans of the carotid arteries, between culprit, non-culprit and asymptomatic carotids. This is compared with the ability of TexRAD (first order) features to distinguish between carotid artery types.

3.1.1 Aims and Objectives

The specific objectives for this chapter are as follows:

Primary objectives:

- To investigate whether there are statistical differences between higher-order radiomic features derived from CT angiography (CTA) and unenhanced CT images of asymptomatic, culprit and non-culprit carotid arteries.
- To evaluate the level of agreement between first order features derived with TexRAD versus PyRadiomics.
- To evaluate the value of multi-slice analysis versus single-slice analysis in carotid artery radiomics.

Secondary objectives:

- To evaluate the utility of resegmentation in identifying differences in radiomic features.
- To evaluate differences in radiomic features stratified by carotid plaque type or carotid stenosis severity.

3.1.2 Hypotheses

- There are statistical differences between the radiomic profiles of asymptomatic, culprit and non-culprit carotid arteries respectively.
- PyRadiomics first order features will be similar to TexRAD first order features and can be used instead for a more extensive radiomic analysis of the carotid arteries.
- Multi-slice analysis reveals more radiomic differences between carotid artery types than single-slice analysis.
- There will be more differences between mixed and noncalcified plaque types, than calcified plaque types.

3.2 Materials and Methods

3.2.1 Carotid CT Datasets

Please see Chapter 2, section 2.2.1 for details regarding carotid CT data acquisition and curation.

3.2.2 Radiomic Analysis with PyRadiomics
3.2.2.1 PyRadiomics Software and Radiomic Feature Classes

This chapter investigated an alternative radiomics platform to TexRAD to enable the extraction of radiomic features, beyond first order histogram features. PyRadiomics is an open-source Python package[173] for radiomic feature extraction from medical images; a detailed description is provided in Chapter 1, section 1.3.2.2. This radiomics package enables the extraction of first order features, including those available in TexRAD, as well as other texture feature classes.

There are 7 different radiomic feature classes available in PyRadiomics (not including other feature classes that can be derived by using additional filters on the medical images), these are: (1) first order, (2) shape, (3) GLCM, (4) GLSZM, (5) GLRLM, (6) NGTDM and (7) GLDM (radiomic feature abbreviations are explained in Table 3.38). Each radiomic feature class contains different numbers of individual radiomic features. The formal definitions and equations for these radiomic features, as calculated by PyRadiomics, are available from the online documentation at https://pyradiomics.readthedocs.io/.

In this study, radiomic features were extracted from all feature classes, except the shape feature class. This is because shape descriptors are independent of the grey-values within the image and instead describes the shape of the ROI which is manually delineated. For analysis of the carotid arteries in this study, the shape descriptors would be a reflection of the carotid artery ROI segmentation rather than capture information about the shape of the carotid plaque for example. Shape descriptors may be more useful in analysing the shape of a tumour[174].

Table 3.38 presents the different feature classes (n=6) generated using PyRadiomics from carotid CTA images, and the number of features from each feature class that were extracted in this study. In total 93 radiomic features were extracted, these are shown in Table 3.39.

Table 3.38 PyRadiomics Feature Classes Studied

Abbreviation	Radiomic Feature Class	No. of Features
IH	First Order Intensity Histogram Statistics	18 (19.4%)
GLCM	Grey Level Co-occurrence Matrix	24 (25.8%)
GLDM	Grey Level Dependence Matrix	14 (15.1%)
GLRLM	Grey Level Run Length Matrix	16 (17.2%)
GLSZM	Grey Level Size Zone Matrix	16 (17.2%)
NGTDM	Neighbouring Grey Tone Difference Matrix	5 (5.4%)
TOTAL		93 (100.0%)

A total of 93 radiomic features were extracted per region-of-interest using PyRadiomics.

Table 3.39 Extracted Features with PyRadiomics

#	Radiomic Feature	#	Radiomic Feature	#	Radiomic Feature
First-Order		32	IMC 1: Information Measures of Correlation	63	Long Run Low Grey Level Emphasis
1	10th Percentile	33	IMC 2: Information Measures of Correlation	64	Low Grey Level Run Emphasis
2	90th Percentile	34	Inverse Variance	65	Run Entropy
3	Energy	35	Joint Average	66	Run Length NonUniformity

4	Entropy	36	Joint Energy	67	Run Length NonUniformity Normalised
5	Interquartile Range	37	Joint Entropy	68	Run Percentage
6	Kurtosis	38	MCC	69	Run Variance
7	Maximum	39	Maximum Probability	70	Short Run Emphasis
8	Mean Absolute Deviation	40	Sum Average	71	Short Run High Grey Level Emphasis
9	Mean	41	Sum Entropy	72	Short Run Low Grey Level Emphasis
10	Median	42	Sum Squares	***GLSZM***	
11	Minimum	***GLDM***		73	Grey Level NonUniformity
12	Range	43	Dependence Entropy	74	Grey Level NonUniformity Normalised
13	Robust Mean Absolute Deviation	44	Dependence NonUniformity	75	Grey Level Variance
14	Root Mean Squared	45	Dependence NonUniformity Normalised	76	High Grey Level Zone Emphasis
15	Skewness	46	Dependence Variance	77	Large Area Emphasis
16	Total Energy	47	Grey Level NonUniformity	78	Large Area High Grey Level Emphasis
17	Uniformity	48	Grey Level Variance	79	Large Area Low Grey Level Emphasis
18	Variance	49	High Grey Level Emphasis	80	Low Grey Level Zone Emphasis
GLCM		50	Large Dependence Emphasis	81	Size Zone NonUniformity
19	Autocorrelation	51	Large Dependence High Grey Level Emphasis	82	Size Zone NonUniformity Normalised
20	Cluster Prominence	52	Large Dependence Low Grey Level Emphasis	83	Small Area Emphasis

21	Cluster Shade	53	Low Grey Level Emphasis	84	Small Area High Grey Level Emphasis
22	Cluster Tendency	54	Small Dependence Emphasis	85	Small Area Low Grey Level Emphasis
23	Contrast	55	Small Dependence High Grey Level Emphasis	86	Zone Entropy
24	Correlation	56	Small Dependence Low Grey Level Emphasis	87	Zone Percentage
25	Difference Average	***GLRLM***		88	Zone Variance
26	Difference Entropy	57	Grey Level NonUniformity	***NGTDM***	
27	Difference Variance	58	Grey Level NonUniformity Normalised	89	Busyness
28	ID	59	Grey Level Variance	90	Coarseness
29	IDM	60	High Grey Level Run Emphasis	91	Complexity
30	IDMN: Inverse Difference Moment Normalised	61	Long Run Emphasis	92	Contrast
31	IDN: Inverse Difference Normalised	62	Long Run High Grey Level Emphasis	93	Strength

GLCM, grey level co-occurrence matrix; GLDM, grey level dependence matrix; GLRLM, grey level run length matrix; GLSZM, grey level size zone matrix; NGTDM, neighbouring grey tone difference matrix

3.2.2.2 Overview of DICOM and NifTI formats

After image acquisition from the CT scanner, CT scans are typically stored in the Digital Imaging and Communications in Medicine (DICOM) format. A DICOM file contains a header with imaging metadata, i.e. information that describes the image, such as the scanning parameters, the image matrix dimensions and sometimes patient information (please note, patient information was been removed from imaging metadata following anonymisation procedures prior to curating the carotid CT dataset). The size of the DICOM header is variable and depends on the image modality and scanner used. A DICOM file also contains the pixel data (the pixel values as integers) itself, therefore by using this format both the metadata and pixel data are merged into a single file.

Pixel values in CT images are measured in Hounsfield units, previously discussed in Chapter 2 section 2.1, however, the pixel values are stored as unsigned integers on disk (i.e. on the hard drive storage space) within the DICOM format. Unsigned integers are whole numbers that can only be positive. However, Hounsfield units can have negative values and so a linear transformation must be applied to the pixel values stored in the DICOM file to calibrate the grey level values within the image to HU. When DICOM images are loaded into memory (random access memory [RAM] of the computer), the Rescale Intercept and Rescale Slope DICOM tags within the metadata is used to specify this linear transformation in the form:

$$U = m \times SV + b$$

Equation 3.4 Linear Transformation from Pixel Values to Hounsfield units

Where U is in Hounsfield units, m is the rescale slope, SV is the stored value and b is the rescale intercept[175].

Another commonly used imaging format was devised by the Neuroimaging Informatics Technology Initiative, referred to as NifTI (or nii format). This was originally developed for brain imaging but is now widely used for most other volume images. This format saves the necessary information to reconstruct the image and orient it in physical space – it stores information such as the data type and voxel spacing within its header of a fixed size. There are several Python libraries that can read .nii files, access the header information and parse it to obtain a reconstructed image as a Numpy array, such as SimpleITK upon which PyRadiomics

is largely based on for image processing needs. NifTI file formats are easier to manage than DICOM in Python and is the preferred file format when using PyRadiomics.

3.2.2.3 Overview of PyRadiomics Requirements

PyRadiomics is compatible with a wide range of medical imaging modalities, including CT, PET and MRI. Requirements for radiomic feature extraction with PyRadiomics include the image being in NifTI format (although there are experimental labs for working directly with DICOM imaging formats), as well as the corresponding ROI being in the form of a binary mask also in NifTI format. The binary mask is aligned with the medical image, in this case the CTA scan, and has the same dimensions as the image with the value of '1' in the areas of the region of interest for radiomic feature extraction, and a '0' everywhere else. Figure 3.35 illustrates the process of binary mask generation in general.

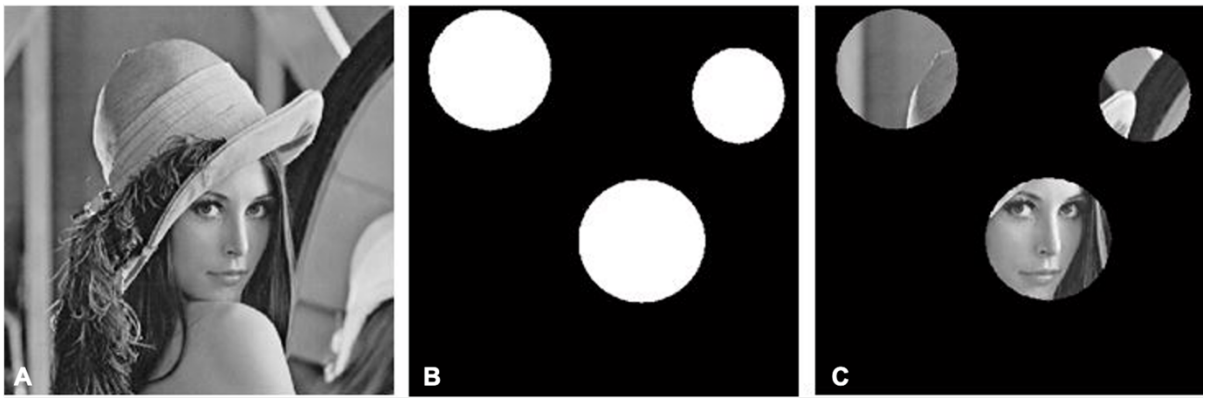


Figure 3.35 Demonstration of how a binary mask works

A) represents the original image, B) is a binary mask which appears dark with a value of '0' in the areas we want to ignore in an image and appears bright with a value of '1' in the regions of interest. C) shows the regions of interest in the image that are selected by the binary mask. This figure was adapted from https://arturomoncadatorres.com/generating-a-binary-mask/

Radiomic feature extraction using PyRadiomics can be executed directly from the command line interface (CLI) with the input image file, the binary mask file and a YAML parameter file (if any) as inputs, and the outputs saved as a .csv file. To batch process images, using a .csv file containing the paths to the respective image files and binary masks as input, can speed up the radiomic workflow. Alternatively, radiomic feature extraction can be performed interactively using a Jupyter Notebook environment.

For this study, PyRadiomics version 3.0 was used for radiomic feature extraction which was executed from the command line. Figure 3.36 presents example code executed from the CLI for the texture feature extraction of the left carotid artery from the carotid CTA of patient 3399 with a parameter file 'resampling.yaml'. A log file 'resampling.log' was produced for debugging purposes, if necessary, and the results were outputted as a .csv file '3399.csv'.

```
~ % pyradiomics 3399_L.nii 3399_L_mask.nii -p resampling.yaml --log-file=resampling.log --logging-level=DEBUG -o 3399.csv -f csv
```

Figure 3.36 Using PyRadiomics from the Command Line Interface

3.2.2.4 DICOM to NifTI conversion

MRI Convert version 2.1.0 build 440, developed by Jolinda Smith and Chuck Theobald from the University of Oregon (2013) converts DICOM files to other medical imaging formats, such as NifTI (.nii). MRI Convert was used to convert individual DICOM slices for single-slice analysis into individual NifTI files with the rescale slope and intercept applied to the data to convert pixel values to Hounsfield units. For multi-slice analysis, 14 consecutive DICOM slices (as per Chapter 2 methodology) were converted into 1 cumulative NifTI image file.

3.2.2.5 Binary Mask Generation

In single-slice analysis: Previously drawn ROIs in TexRAD were saved as XML files (Chapter 2, section 2.2.4.3), which store the points of each ROI in the form of pixel coordinates (x, y). To use the same ROIs in PyRadiomics for direct comparability, the x, y coordinates of the ROI were extracted from the XML file using the Python module Element Tree in a Jupyter notebook and converted into a 512x512 voxel binary mask and saved as a .nii file using the modules scikit-image and SimpleITK in Python 3.7.4. The process is illustrated in Figure 3.37.

A)

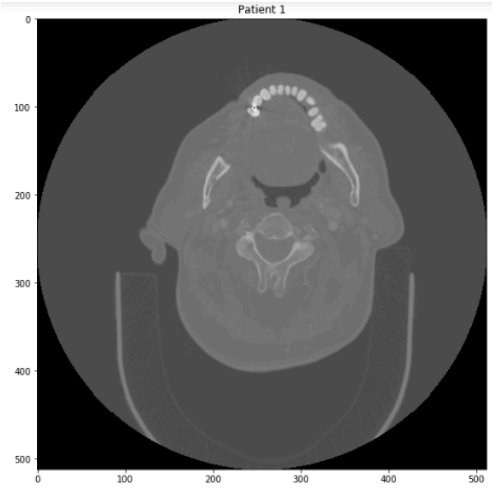
```
import xml.etree.ElementTree as ET
tree = ET.parse(roi_path)
root = tree.getroot()
```

```
x = []
y = []

for i in root.iter('X'):
    x.append(float(i.text))

for i in root.iter('Y'):
    y.append(float(i.text))
coordinates = list(zip(x,y))
```

```
img=Image.new('L', (512, 512), 0)
ImageDraw.Draw(img).polygon(coordinates, outline=1, fill=1)
mask=numpy.array(img)
```

```
plt.imshow(mask)
```

B)



C)

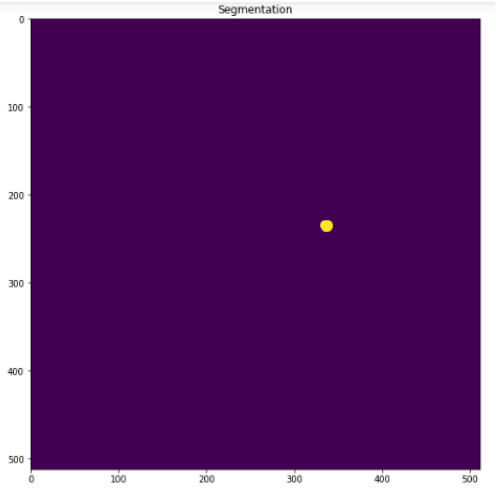


Figure 3.37 Extracting ROI coordinates form XML file and Binary Mask creation

A) ElementTree is a Python module that can extract information from XML files, this is used to extract the x, y coordinates of the ROI to create a C) binary mask as a numpy array that is subsequently saved as a .nii file for compatibility of use with PyRadiomics. B) represents the axial CT angiography slice to be analysed.

In multi-slice analysis: Since ROIs in TexRAD can only be saved for individual slices, there were 14 individual XML files per carotid artery in multi-slice analysis. These individual ROIs had to be combined together for multi-slice analysis with PyRadiomics in Nifti format. To do this, binary masks for each individual XML file were generated and saved as individual NifTI files, as per above. The 14 individual binary masks were subsequently converted into a Numpy array, then vertically stacked together into a single 3D volume and then converted into a single Nifti file. Therefore, for each carotid artery, the 14 consecutive slices became a single VOI from which 3D radiomic features could be extracted.

To visualise the multi-slice ROIs, the marching cubes algorithm implemented in the Python scikit-image module was used to generate a 3D isosurface using triangulations, as shown in Figure 3.38 panel C. Marching cubes is an algorithm to extract a 2D surface mesh from a 3D volume[176].

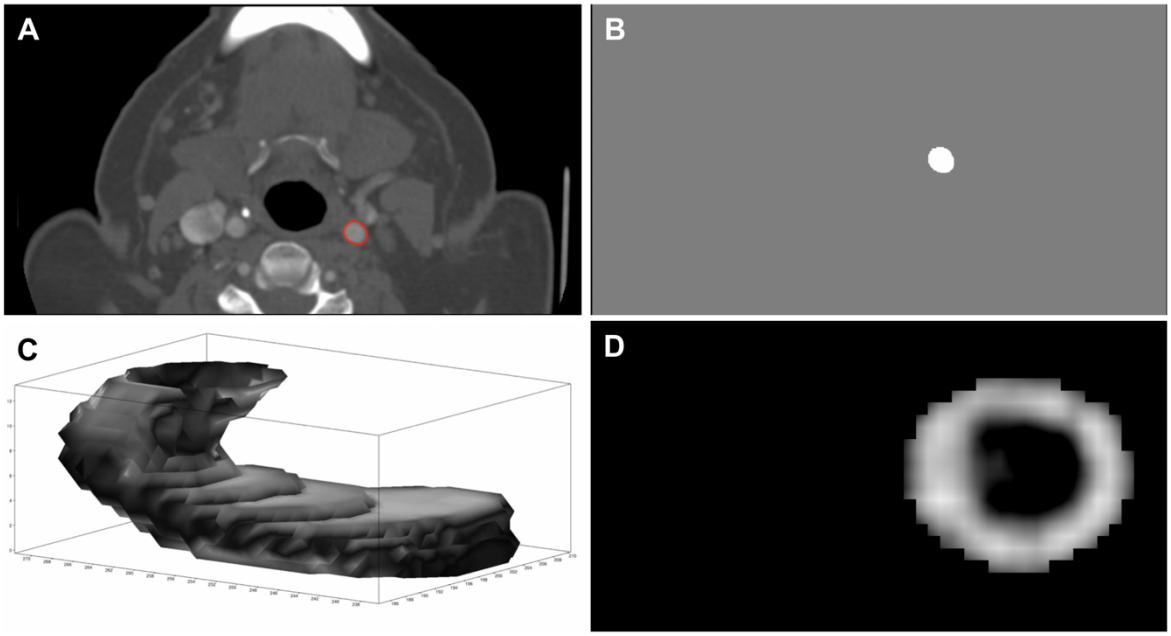


Figure 3.38 PyRadiomics ROI mask generation and feature extraction

A. ROI segmentation of Carotid CTA
B. Mask generation in NifTI format
C. 3D visualisation (marching cubes algorithm) of volume-of-interest
D. Voxel-based extraction feature map (PyRadiomics) of GLCM Joint Entropy (for visualisation purposes only)

This figure was reproduced from the abstract by Le et al. 2020[177].

3.2.2.6 PyRadiomics Image Pre-processing Considerations

There are several image pre-processing considerations prior to radiomic feature extraction with PyRadiomics. Specific parameters associated with these considerations can be specified using YAML files that are supplied as inputs along with the NifTI image, and the binary mask to PyRadiomics. In this study, two different image pre-processing schemes were investigated: (a) Original image (no image pre-processing) and (b) Resegmentation (which is equivalent to the Restricted algorithms used in TexRAD for comparability).

Resegmentation: Similar to TexRAD's ability to implement different algorithms that analysed pixels within different ranges of CT Hounsfield units. Resegmentation refers to the same process whereby the upper and lower limits of permitted pixel values for consideration in radiomic feature extraction, are specified[162].

In PyRadiomics, if resegmentation was applied, the values of the TexRAD Restricted I algorithm were used: 0-200 HUs. This excluded the effects of excess macrocalcification and limited the effect of luminal contrast in the ROI and perivascular fat that might have been captured in the ROI due to human error in segmentation.

YAML files: To specify different image pre-processing considerations in the PyRadiomics workflow, the parameters need to be specified in a YAML file, as shown in Figure 3.39.

```yaml
imageType:
  Original: {}

setting:
  binWidth: 25
  label: 1
  resegmentRange: [0, 200]

featureClass:
  firstorder:
  glcm:
  gldm:
  glrlm:
  glszm:
  ngtdm:
```

Figure 3.39 Example YAML file with PyRadiomics Configurations

In this YAML files, imageType refers to which type of image(s) are used for radiomic feature extraction. 'Original' refers to the image without filtration, there are several options available in PyRadiomics for prior filtration to be performed before texture feature extraction, such as a Gaussian filter. The bin width is fixed to a size of 25, the value of '1' indicates the values of the ROI in the binary mask. Resegmentation is applied to analyse only the pixel values between 0 to 200 HUs. Features from 6 features classes are extracted: first order statistics and those derived from the GLCM, GLDM, GLRLM, GLSZM and NGTDM matrices

PyRadiomics 2D single-slice analysis: certain radiomic features require isotropic images (images that have the same pixel size in all directions). For single-slice analysis, this requirement is only in the x and y directions of the axial slice. All CT and CTA images had in-plane isotropy of between 0.4x0.4mm to 0.6x0.6mm, therefore no resampling was required for single-slice radiomic feature extraction. Radiomic features were therefore extracted from the original 2D slices without prior image resampling.

PyRadiomics 3D multi-slice analysis: when extracting radiomic features from the 3D volume, the images must be isotropic in all axes (*x*, *y* and *z* planes have identical pixel sizes), particularly for the calculation of higher-order radiomic features to be rotationally invariant as stated in the image biomarker standardisation initiative documentation[178] and as recommended by PyRadiomics documentation. In CT imaging, images are often isotropic in-plane (*x* and *y* are equal) but will have a larger *z*-axis slice spacing and therefore be anisotropic in 3D.

Therefore, in radiomic studies, it is common for images to be isotropically resampled[169]. In this study, the images and binary masks were resampled to 1x1x1mm pixel spacing when using the multi-slice approach. PyRadiomics uses the Python module SimpleITK to interpolate images, and in this study cubic B-spline interpolation was used as this was the PyRadiomics default setting. We investigate the impact of using different interpolation methods in Chapter 4.

3.2.2.7 Image Quantisation

Image quantisation refers to the conversion of image grey level values to a discrete set of grey level counts. Before radiomic feature extraction, the image must be quantised either by using a fixed number of bins (BN), or by using a fixed bin width (BW; also referred to as bin size). The default settings in PyRadiomics was used in this study: a fixed bin width of 25. Figure 3.40 demonstrates the image quantisation process. We investigate the impact of using different image quantisation methods and values in Chapter 4.

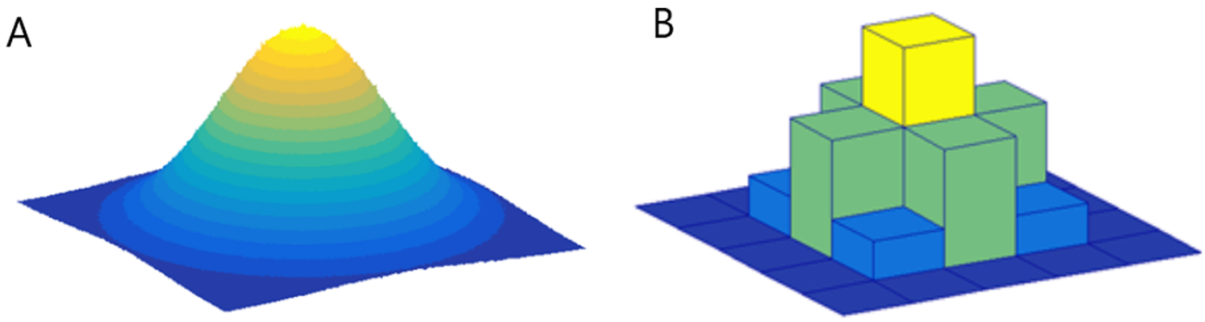


Figure 3.40 Image Intensity Quantisation

A) Original data, B) a generic discretised version. Figure reproduced from van Timmerman et al. 2020[73]

3.2.2.8 PyRadiomics Subset Analysis

3.2.2.8.1 Comparisons with asymptomatic carotid plaque only

As in Chapter 2, section 2.2.4.5 with TexRAD subset analysis, a similar methodology was employed for PyRadiomics features. To explore whether the presence or absence of visible carotid atheroma in asymptomatic carotid arteries impacted the PyRadiomics feature differences with culprit or non-culprit carotid arteries respectively, two levels of analyses were conducted: (1) inter-patient comparisons using all asymptomatic carotid arteries (50 carotid arteries) and (2) inter-patient comparisons using only asymptomatic carotid arteries with visible plaque (28 carotid arteries) versus 41 culprit and 41 non-culprit carotid arteries respectively. These analyses were conducted for both single-slice and multi-slice approaches.

3.2.2.8.2 Carotid Plaque Type and Carotid Stenosis Severity

A subset analysis in the multi-slice approach was conducted to explore PyRadiomics features stratified by carotid plaque type (calcified, noncalcified and mixed) as described in Chapter 2, section 2.2.4.5.2. In addition, subset analysis of PyRadiomics features was conducted for two categories of carotid stenosis severity: below 50% and above 50%, please see Chapter 2, section 2.2.4.5.3 for further details.

3.2.3 Unenhanced CT Analysis

ROIs drawn for the CTA images were used to analyse unenhanced CT scans as described in Chapter 2, section 2.2.5. The ROIs were processed for radiomic feature extraction using PyRadiomics as described above and created into VOIs for multi-slice analysis.

Two settings were investigated: (1) multi-slice analysis of the original image resampled to 1x1x1mm with B-spline interpolation and fixed bin width of 25 and (2) multi-slice analysis of image with resegmentation [0, 200] HU resampled to 1x1x1 mm with B-spline interpolation and fixed bin width of 25.

3.2.4 Statistical Analysis

Reporting of descriptive statistics and statistical comparisons between culprit versus non-culprit carotid arteries within symptomatic patients, as well as those between asymptomatic and symptomatic patients followed the methodology specified in Chapter 2, section 2.2.6.

To measure the level of agreement between first order features derived with TexRAD versus those from PyRadiomics, the two-way mixed effects model the intraclass correlation coefficient (ICC) for absolute agreement and consistency were calculated, with the single measures value reported[167].

All statistical analysis was performed in IBM SPSS Statistics for Macintosh (Version 25.0. Armonk, NY, USA: IBM Corp.), Microsoft Excel for Mac (Version 16.37), R 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) and Python 3.7.4.

3.3 Results

3.3.1 TexRAD versus PyRadiomics

ICC values were calculated in terms of absolute agreement and consistency to assess the concordance between first order texture features calculated by TexRAD versus those calculated by PyRadiomics. TexRAD calculates six first order texture features which describe the histogram of grey level values within an ROI, see Table 3.40.

In PyRadiomics, a greater range of first order features are calculated such as First Order: 10^{th} percentile and First Order: 90^{th} percentile, compared with those available from TexRAD. PyRadiomics does calculate the following TexRAD texture features: (1) First Order: Mean, (2) First Order: Entropy, (3) First Order: Skewness and (4) First Order: Kurtosis. The texture feature 'mean of positive pixels' however is not calculated in PyRadiomics, and rather than First Order: Standard Deviation, PyRadiomics calculates the texture feature called First Order: Variance as default. We can convert First Order: Variance to First Order: Standard Deviation by squaring rooting the values of the former texture feature.

Table 3.40 Intraclass Correlation Coefficients between TexRAD and PyRadiomics first order CT angiography radiomic features

Single-slice Analysis	ICC absolute agreement	ICC consistency
Mean	0.838 (0.662, 0.913)	0.872 (0.808 0.916)
MPP	-	-
Standard deviation	0.639 (0.294, 0.803)	0.722 (0.599, 0.811)
Entropy	0.122 (-0.073, 0.358)	0.379 (0.178, 0.550)
Skewness	0.844 (0.529, 0.931)	0.896 (0.844, 0.932)
Kurtosis	0.571 (-0.066, 0.857)	0.897 (0.845, 0.933)

Single measures ICC and 95% confidence intervals. reported; ICCs calculated using culprit and non-culprit carotid single-slice analysis texture feature values with no prior filtering or resegmentation. For PyRadiomics features, the default setting of 25 was used for the bin width in image quantisation.

There were moderate-to-good levels of agreement (ICC >0.5) between the TexRAD and PyRadiomics regarding the following features: (1) First Order: Mean, (2) First Order: Standard Deviation, (3) First Order: Skewness and (4) First Order: Kurtosis. There was poor agreement for First Order: Entropy (ICC<0.5).

To investigate the cause for the discrepancy between First Order: Entropy as calculated in TexRAD versus PyRadiomics, the team at Feedback Medical Ltd (the company that produces TexRAD software) were contacted. Antoine Saillant, a data scientist at Feedback Medical Ltd kindly compared the unfiltered texture features of PyRadiomics with TexRAD features and found the following differences:

- First Order: Mean: there is a systematic difference of 0.5 between the two implementations because of differences in rounding.
- First Order: Kurtosis: there is a systematic difference of 3 between the two implementations due to the definitions of kurtosis used in TexRAD versus PyRadiomics.
- First Order: Entropy: there is a ratio of the natural logarithm of 2 (ln (2)) between the two features because PyRadiomics calculates the entropy with a log base 2, whilst TexRAD uses the natural logarithm (ln). In addition, TexRAD histograms use a bin width of 1, whereas the default of PyRadiomics is a bin width of 25. This difference in image quantisation value and entropy definition gives rise to the discrepancy in ICC values.

As these differences, which are explainable by the use of different texture feature definitions, were systematic, the statistics and correlations previously described in Chapter 2 using TexRAD, should be similar when using PyRadiomics going forward.

3.3.2 Single-Slice Approach

3.3.2.1 Original Image (TexRAD unrestricted algorithm equivalent)

In single-slice analysis, an ROI was drawn on the axial slice best representing the carotid bifurcation per carotid artery. When comparing culprit versus non-culprit carotid arteries, this involved paired analysis within the symptomatic patients. There were 41 symptomatic patients and therefore 82 carotid arteries (41 were culprit and 41 were non-culprit carotid arteries).

PyRadiomics was used to extract 93 radiomic features from 6 different radiomic feature classes from each ROI. To be comparable with the approaches used in TexRAD using the unrestricted and restricted algorithms (see Chapter 2, section 2.2.4.1), PyRadiomics was used to extract radiomic features from the original image (equivalent to TexRAD Unrestricted algorithm) and following resegmentation [0, 200] HU (equivalent to TexRAD Restricted I algorithm).

3.3.2.1.1 Culprit versus Non-culprit Carotid Arteries

Following paired analysis between culprit versus non-culprit carotid arteries, 5 radiomic features had a p-value<0.05. These are summarised in Table 3.41. These features were no longer statistically significant following Benjamini-Hochberg correction for multiple comparisons.

Table 3.41 Single-slice Analysis: Original Image - Culprit versus Non-culprit

Radiomic Feature	Culprit Carotids (n=41)	Non-culprit Carotids (n=41)	P-value
First Order: Median	198.5 (132.0 – 303.5)	234.5 (193.0 – 325.0)	0.027
GLCM: Maximum Probability	0.065 (0.047 – 0.082)	0.079 (0.051 – 0.095)	0.036
GLDM: Dependence Variance	2.69 (1.94 – 3.02)	2.87 (2.35 – 3.78)	0.013
GLDM: Large Dependence High Grey Level Emphasis	1439.4 (1008.6 – 2076.8)	1998 (1311-2597)	0.002
GLSZM: Large Area High Grey Level Emphasis	2667.7 (1479.2 – 4814.8)	4237.8 (1771.0 – 6675.8)	0.017

Median and IQR reported. The p-value was obtained using Wilcoxon signed-rank test. The table reports the values for the radiomic features that had a p<0.05 when compared between culprit versus non-culprit carotid arteries.

3.3.2.1.2 Asymptomatic versus Symptomatic Patients

When comparing the carotid arteries of asymptomatic patients with those of symptomatic patients, this was conducted on a per vessel type basis i.e. (1) asymptomatic carotid versus culprit carotid artery and (2) asymptomatic carotid versus non-culprit carotid artery. Without resegmentation, 51 out of the 93 (54.8%) radiomic features had a p-value <0.05 when comparing between Asx versus CC carotid arteries, and 50 out of 93 (53.8%) had a p-value <0.05 when comparing Asx versus NC carotid arteries.

Following, Benjamini-Hochberg correction, 41 radiomic features remained statistically significantly different between asymptomatic versus culprit carotid arteries; and 35 radiomic features were different between asymptomatic versus non-culprit carotid arteries. Therefore, there were less differences between asymptomatic versus non-culprit carotid arteries than between asymptomatic versus culprit carotid arteries. Differences in the radiomic features spanned across all 6 feature classes, see Table 3.42.

Table 3.42 Single-slice Analysis: Original Image - Asymptomatic versus Symptomatic

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC		
First Order (n=18)	9 (50.0%)	8 (44.4%)
GLCM (n=24)	19 (79.2%)	11 (45.8%)
GLDM (n=14)	9 (64.3%)	9 (64.3%)
GLRLM (n=16)	7 (43.8%)	6 (37.5%)
GLSZM (n=16)	5 (31.3%)	5 (31.3%)
NGTDM (n=5)	2 (40.0%)	2 (40.0%)
Total (n=93)	51 (54.8%)	41 (44.1%)
Asx versus NC		
First Order (n=18)	9 (50.0%)	8 (44.4%)
GLCM (n=24)	16 (66.7%)	10 (41.7%)
GLDM (n=14)	8 (57.1%)	6 (42.9%)
GLRLM (n=16)	8 (50.0%)	4 (25.0%)
GLSZM (n=16)	7 (43.8%)	5 (31.3%)
NGTDM (n=5)	2 (40.0%)	2 (40.0%)
Total (n=93)	50 (53.8%)	35 (37.6%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery.

3.3.2.2 Resegmented Image (TexRAD restricted I algorithm equivalent)
3.3.2.2.1 Culprit versus Non-culprit Carotid Arteries

Following resegmentation, there were 22 radiomic features with a p-value<0.05 when comparing culprit versus non-culprit carotids, spanning 5 feature classes. These are shown in Table 3.43, however, they were no longer statistically significant following Benjamini-Hochberg correction.

Table 3.43 Single-slice Analysis: Resegmentation - Culprit versus Non-culprit

Feature Class	Feature Name	Culprit Carotids (n=41)	Non-culprit Carotids (n=41)	P-value
First Order	90th Percentile	165.9 (158.0. – 173.4)	172.5 (166.0-179.4)	0.007
	Mean	97.5 (86.2-105.1)	100.5 (93.7-113.7)	0.019
	Median	89.5 (80.0-99.5)	95.0 (86.5-115.0)	0.025
	Root Mean Squared	106.0 (98.6-115.4)	113.4 (104.9-124.0)	0.013
	Skewness	0.334 (0.137-0.534)	0.202 (-0.058-0.400)	0.015
GLCM	Autocorrelation	18.8 (16.6-21.4)	21.0 (18.0-25.7)	0.004
	Cluster Shade	6.96 (1.27-9.76)	2.49 (-4.08-7.67)	0.026
	Contrast	3.23 (2.66-4.25)	4.06 (3.27-4.54)	0.038
	Difference Entropy	2.09 (1.94-2.21)	2.15 (2.06-2.25)	0.010
	Difference Variance	1.37 (1.11-1.75)	1.63 (1.35-1.84)	0.025
	Inverse Variance	0.455 (0.427-0.471)	0.422 (0.393-0.448)	0.002
	Joint Average	4.25 (3.93-4.50)	4.44 (4.10-4.91)	0.004
	Sum Average	8.50 (7.85-9.00)	8.89 (8.19-9.81)	0.004
GLDM	High Grey Level Emphasis	22.3 (19.3-24.6)	24.7 (21.2-29.1)	0.003
	Large Dependence Low Grey Level Emphasis	0.971 (0.638-1.365)	0.669 (0.464-1.091)	0.036
	Small Dependence High Grey Level Emphasis	10.9 (8.62-12.7)	11.6 (10.1-14.0)	0.017
GLRLM	High Grey Level Run Emphasis	22.9 (20.6-25.7)	25.4 (22.5-29.5)	0.003
	Short Run High Grey Level Emphasis	20.8 (18.4-23.4)	23.4 (20.6-26.6)	0.003
GLSZM	High Grey Level Zone Emphasis	26.3 (23.5-30.0)	27.7 (25.4-30.2)	0.029
	Large Area Low Grey Level Emphasis	1.12 (0.618-2.74)	0.852 (0.456-1.57)	0.013
	Size Zone NonUniformity	26.4 (23.8-33.9)	32.5 (24.6-40.1)	0.050
	Small Area High Grey Leve Emphasis	18.6 (15.9-21.3)	19.3 (17.9-22.6)	0.041

Median and IQR reported. The p-value was obtained using Wilcoxon signed-rank test.

3.3.2.2.2 Asymptomatic versus Symptomatic Patients

There were more differences between asymptomatic and culprit carotid arteries (64.5% of radiomic features following Benjamini-Hochberg correction), than between asymptomatic and non-culprit carotid arteries (38.7%). Resegmentation increased the number of differences between the carotid arteries of asymptomatic patients versus symptomatic patients, versus no resegmentation as in section 3.3.2.2.1. Asymptomatic carotid arteries differed the most in their GLCM matrix-derived radiomic features compared with culprit carotid arteries, with 78.6% of the GLCM features being different. The radiomic feature class with the least differences between Asx and CC carotid arteries were those derived from the NGTDM class (40.0%). In contrast, asymptomatic carotid arteries differed the most in the GLDM feature class with non-culprit carotid arteries (64.3%) and were quite similar to non-culprit carotid arteries in terms of first order radiomic features (differing in only 1 first order radiomic feature following resegmentation).

Table 3.44 Single-slice Analysis: Resegmentation - Asymptomatic versus Symptomatic

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC: Single-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	11 (61.1%)	10 (55.6%)
GLCM (n=24)	18 (75.0%)	17 (70.8%)
GLDM (n=14)	11 (78.6%)	11 (78.6%)
GLRLM (n=16)	10 (62.5%)	10 (62.5%)
GLSZM (n=16)	10 (62.5%)	10 (62.5%)
NGTDM (n=5)	2 (40.0%)	2 (40.0%)
Total (n=93)	62 (66.7%)	60 (64.5%)
Asx versus NC: Single-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	1 (5.56%)	1 (5.56%)
GLCM (n=24)	10 (41.7%)	10 (41.7%)
GLDM (n=14)	10 (71.4%)	9 (64.3%)
GLRLM (n=16)	9 (56.3%)	7 (43.8%)
GLSZM (n=16)	8 (50.0%)	8 (50.0%)
NGTDM (n=5)	2 (40.0%)	1 (20.0%)
Total (n=93)	40 (43.0%)	36 (38.7%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

3.3.2.3 Single-slice analysis: asymptomatic with carotid atherosclerosis only

Table 3.45 Single-slice analysis: Asx with plaque only

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC: Single-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	4 (22.2%)	3 (16.7%)
GLCM (n=24)	5 (20.8%)	3 (12.5%)
GLDM (n=14)	6 (42.9%)	4 (28.6%)
GLRLM (n=16)	4 (25.0%)	4 (25.0%)
GLSZM (n=16)	4 (25.0%)	3 (18.8%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	23 (24.7%)	17 (18.3%)
Asx versus NC: Single-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	2 (11.1%)	2 (11.1%)
GLCM (n=24)	3 (12.5%)	3 (12.5%)
GLDM (n=14)	3 (21.4%)	3 (21.4%)
GLRLM (n=16)	4 (25.0%)	2 (12.5%)
GLSZM (n=16)	4 (25.0%)	1 (6.25%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	16 (17.2%)	11 (11.8%)
Asx versus CC: Single-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	7 (38.9%)	2 (11.1%)
GLCM (n=24)	11 (45.8%)	10 (41.7%)
GLDM (n=14)	11 (78.6%)	10 (71.4%)
GLRLM (n=16)	9 (56.3%)	8 (50.0%)
GLSZM (n=16)	9 (56.3%)	9 (56.3%)
NGTDM (n=5)	2 (40.0%)	2 (40.0%)
Total (n=93)	49 (52.7%)	41 (44.1%)
Asx versus NC: Single-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	2 (11.1%)	1 (5.56%)
GLCM (n=24)	7 (29.2%)	7 (29.2%)
GLDM (n=14)	8 (57.1%)	7 (50.0%)
GLRLM (n=16)	8 (50.0%)	7 (43.8%)
GLSZM (n=16)	7 (43.8%)	7 (43.8%)
NGTDM (n=5)	1 (20.0%)	0 (0.0%)
Total (n=93)	33 (35.5%)	29 (31.2%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

With the unrestricted algorithm, considering Asx carotid arteries with plaque only resulted in less differences with CC and NC carotids respectively compared with considering all Asx carotids. Previously, there were 41 features that were statistically significantly different between CC and Asx, and 35 between NC and Asx (see Table 3.42). This decreased to 17 for CC vs Asx and 11 for NC vs Asx (see Table 3.45).

Following resegmentation, when considering only asymptomatic carotid arteries with visible carotid plaque, there were less differences between CC vs Asx and NC vs Asx than when all asymptomatic carotid arteries were included in the analysis. When considering all Asx carotids, 60 radiomic features were statistically significantly different between CC and Asx and 36 features were different between NC and Asx carotid arteries (even after correction for multiple testing). However, this decreased to 41 between CC and Asx and 29 between NC and Asx when considering Asx with plaque only.

3.3.3 Multi-slice Approach

3.3.3.1 Original Image (TexRAD unrestricted algorithm equivalent)
3.3.3.1.1 Culprit versus Non-culprit Carotid Arteries

When comparing culprit versus non-culprit carotid arteries without resegmentation, there were 14 radiomic features with p-value<0.05 and 3 remained statistically significantly different after Benjamini-Hochberg correction, shown in Table 3.46.

Table 3.46 Multi-slice Analysis: Original - Culprit versus Non-culprit

Feature Class	Feature Name	Culprit Carotids (n=41)	Non-culprit Carotids (n=41)	P-value
First Order	10th Percentile	40.8 (26.1-49.7)	44.8 (31.3-57.5)	0.015*
	Mean	182.3 (158.0-250.9)	219.0 (184.5-277.2)	0.006*
	Median	140.0 (107.5-212.8)	203.0 (155.4-262.6)	0.0001**
GLCM	Joint Average	12.4 (10.0-16.0)	14.3 (11.7-17.4)	0.025*
	Maximum Probability	0.044 (0.030-0.056)	0.050 (0.040-0.065)	0.038*
	Sum Average	24.8 (20.1-32.0)	28.7 (23.3-34.8)	0.025*
GLDM	Large Dependence High Grey Level Emphasis	6044 (4174-7713)	8214 (6372-9891)	0.001**
	Large Dependence Low Grey Level Emphasis	0.519 (0.300-0.842)	0.319 (0.224-0.601)	0.003*
	Low Grey Level Emphasis	0.020 (0.015-0.023)	0.018 (0.013-0.022)	0.026*
GLRLM	Long Run High Grey Level Emphasis	300.1 (223.1-425.9)	363.1 (282.4-454.2)	0.016*
	Long Run Low Grey Level Emphasis	0.030 (0.023-0.040)	0.027 (0.019-0.032)	0.012*

GLSZM	Grey Level NonUniformity	23.6 (19.7-25.8)	24.1 (21.4-28.1)	0.009*
	Large Area High Grey Level Emphasis	30717 (13331-52606)	52776 (25325-100329)	0.0002**
	Large Area Low Grey Level Emphasis	4.88 (0.701-10.4)	1.42 (0.582-4.37)	0.030*

*Median and IQR reported. The p-value was obtained using Wilcoxon signed-rank test, * p-value <0.05; ** statistically significant even after Benjamini-Hochberg correction.*

3.3.3.1.2 Asymptomatic versus Symptomatic Patients

Even without resegmentation, using a multi-slice approach resulted in more differences between carotid arteries of asymptomatic patients and those of symptomatic patients than a single-slice approach (multi-slice: Asx versus CC: 48.4% versus single-slice: 44.1%), see Table 3.47. There were differences between the different carotid artery types across all radiomic feature classes.

Table 3.47 Multi-slice Analysis: Original - Asymptomatic versus Symptomatic

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC		
First Order (n=18)	7 (38.9%)	7 (38.9%)
GLCM (n=24)	17 (70.8%)	15 (62.5%)
GLDM (n=14)	6 (42.9%)	6 (42.9%)
GLRLM (n=16)	7 (43.8%)	7 (43.8%)
GLSZM (n=16)	8 (50.0%)	7 (43.8%)
NGTDM (n=5)	5 (100.0%)	3 (60.0%)
Total (n=93)	50 (53.8%)	45 (48.4%)
Asx versus NC		
First Order (n=18)	10 (55.6%)	8 (44.4%)
GLCM (n=24)	16 (66.7%)	14 (58.3%)
GLDM (n=14)	7 (50.0%)	5 (35.7%)
GLRLM (n=16)	7 (43.8%)	4 (25.0%)
GLSZM (n=16)	7 (43.8%)	7 (43.8%)
NGTDM (n=5)	4 (80.0%)	4 (80.0%)
Total (n=93)	51 (54.8%)	42 (45.2%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

In multi-slice analysis, without resegmentation, the greatest proportion of differences for Asx versus CC was in the GLCM feature class. For Asx versus NC, this was the NGTDM feature class. There remain a greater number of differences between Asx versus CC carotid arteries than between Asx versus NC carotid arteries.

3.3.3.2 Resegmented Image (TexRAD restricted I algorithm equivalent)
3.3.3.2.1 Culprit versus Non-culprit Carotid Arteries

The multi-slice approach with resegmentation greatly increased the number of differences in radiomic features between culprit versus non-culprit carotid arteries compared with (1) multi-slice analysis with no resegmentation, (2) single-slice analysis with resegmentation and (3) single-slice analysis with no resegmentation.

With resegmentation, in the multi-slice approach, there were 57 radiomic features with p-value<0.05 when comparing culprit versus non-culprit carotid arteries. There were 54 radiomic features remaining statistically significantly different after Benjamini-Hochberg correction, these are shown in Table 3.48. The differences spanned all 6 feature classes examined.

Table 3.48 Multi-slice Analysis: Resegmentation - Culprit versus Non-culprit

Feature Class	Feature Name	Culprit Carotids (n=41)	Non-culprit Carotids (n=41)	P-value
First Order	90th Percentile	166.0 (159.9-171.7)	173.4 (167.7-179.6)	<0.0001
	Interquartile Range	68.4 (61.5-80.1)	76.5 (68.5-82.1)	0.014
	Kurtosis	2.40 (2.16-2.62)	2.20 (2.05-2.35)	0.003
	Mean	91.5 (86.3-99.7)	97.2 (89.1-105.8)	0.001
	Mean Absolute Deviation	39.7 (37.4-43.1)	41.4 (39.2-44.1)	0.029
	Median	85.0 (76.4-93.9)	92.4 (80.2-101.1)	0.003
	Robust Mean Absolute Deviation	28.9 (26.6-32.7)	30.9 (28.7-33.5)	0.013
	Root Mean Squared	102.8 (97.5-110.5)	109.1 (102.1-117.2)	0.0003
	Skewness	0.445 (0.219-0.648)	0.241 (0.095-0.480)	0.0004
GLCM	Autocorrelation	17.4 (16.3-19.4)	19.2 (16.5-22.5)	0.001
	Cluster Shade	8.95 (3.97-11.8)	6.17 (2.27-9.51)	0.007
	Contrast	4.72 (4.23-5.82)	5.66 (5.25-6.26)	0.008
	Correlation	0.265 (0.214-0.340)	0.232 (0.173-0.276)	0.013
	Difference Average	1.62 (1.49-1.83)	1.80 (1.73-1.95)	0.006
	Difference Entropy	2.34 (2.26-2.44)	2.44 (2.39-2.50)	0.004
	Difference Variance	2.08 (1.86-2.31)	2.32 (2.16-2.58)	0.012
	Id	0.513 (0.494-0.539)	0.485 (0.468-0.502)	0.006
	Idm	0.453 (0.429-0.486)	0.422 (0.396-0.439)	0.005

	Idmn	0.940 (0.928-0.945)	0.930 (0.922-0.933)	0.006
	Idn	0.849 (0.833-0.859)	0.834 (0.823-0.840)	0.005
	Inverse Variance	0.406 (0.386-0.423)	0.380 (0.365-0.395)	0.001
	Joint Average	4.08 (3.89-4.30)	4.26 (3.99-4.62)	0.001
	Joint Energy	0.032 (0.028-0.037)	0.028 (0.024-0.032)	0.021
	Maximum Probability	0.075 (0.065-0.098)	0.067 (0.053-0.076)	0.016
	Sum Average	8.15 (7.78-8.61)	8.52 (7.97-9.25)	0.001
GLDM	Dependence Entropy	5.74 (5.35-5.96)	5.53 (5.32-5.71)	0.003
	Dependence NonUniformity Normalised	0.123 (0.105-0.173)	0.157 (0.142-0.200)	0.001
	Dependence Variance	7.93 (3.90-10.6)	4.15 (3.04-6.09)	0.006
	Grey Level NonUniformity	174.2 (102.7-222.2)	130.8 (106.8-163.3)	0.002
	High Grey Level Emphasis	20.7 (18.8-23.7)	23.2 (20.3-26.3)	0.0002
	Large Dependence Emphasis	27.7 (14.5-38.0)	17.6 (11.9-22.7)	0.003
	Large Dependence Low Grey Level Emphasis	3.60 (2.09-6.07)	2.35 (1.54-3.54)	0.001
	Small Dependence Emphasis	0.173 (0.147-0.268)	0.233 (0.184-0.287)	0.003
	Small Dependence High Grey Level Emphasis	5.01 (4.09-7.65)	6.70 (4.98-8.27)	0.001
	Small Dependence Low Grey Level Emphasis	0.026 (0.019-0.032)	0.032 (0.024-0.036)	0.027
GLRLM	Grey Level NonUniformity	139.8 (92.4-178.4)	117.3 (92.3-142.1)	0.004
	High Grey Level Run Emphasis	21.4 (20.0-24.5)	23.6 (21.1-26.6)	0.0003
	Long Run Emphasis	1.58 (1.34-1.79)	1.40 (1.31-1.52)	0.003
	Long Run High Grey Level Emphasis	29.8 (27.8-32.8)	30.5 (28.1-34.5)	0.013
	Long Run Low Grey Level Emphasis	0.226 (0.189-0.283)	0.199 (0.158-0.257)	0.017
	Run Entropy	3.44 (3.31-3.56)	3.37 (3.31-3.47)	0.005
	Run Length NonUniformity Normalised	0.773 (0.734-0.848)	0.828 (0.800-0.868)	0.001
	Run Percentage	0.866 (0.830-0.913)	0.900 (0.881-0.895)	0.002
	Run Variance	0.221 (0.131-0.317)	0.150 (0.117-0.194)	0.005

	Short Run Emphasis	0.899 (0.879-0.935)	0.925 (0.911-0.944)	0.001
	Short Run High Grey Level Emphasis	20.0 (18.2-22.4)	22.5 (19.8-25.2)	0.0002
GLSZM	Large Area Emphasis	375.9 (38.5-740.1)	65.5 (24.2-168.9)	0.004
	Large Area High Grey Level Emphasis	4122.7 (514.1-7399.3)	884.7 (353.2-2858.7)	0.004
	Large Area Low Grey Level Emphasis	38.4 (3.92-83.0)	7.42 (3.35-23.6)	0.0005
	Zone Percentage	0.197 (0.158-0.333)	0.278 (0.212-0.360)	0.003
	Zone Variance	346.5 (29.5-690.9)	48.9 (16.5-146.3)	0.003
NGTDM	Complexity	26.7 (24.2-29.7)	29.7 (28.1-32.3)	0.0006
	Contrast	0.158 (0.125-0.190)	0.190 (0.163-0.213)	0.019
	Strength	0.117 (0.102-0.142)	0.118 (0.107-0.142)	0.028

Table 3.49 Multi-slice Analysis: Resegmentation - Features with p<0.05 but not statistically significant following corrections

Radiomic Feature	Culprit Carotids	Non-culprit Carotids	P-value
First Order: Uniformity	0.152 (0.140-0.165)	0.144 (0.139-0.154)	0.0006
GLCM: Joint Entropy	5.41 (5.21-5.51)	5.52 (5.40-5.67)	0.019
GLSZM: Grey Level NonUniformity	29.8 (26.2-35.0)	31.9 (26.8-37.9)	0.028

These features in this table were no longer considered statistically significant following Benjamini-Hochberg correction for multiple testing.

3.3.2.2.2 Asymptomatic versus Symptomatic Patients

Multi-slice analysis with resegmentation led to the highest number of differences out of all the different settings investigated when comparing the carotid arteries from asymptomatic versus symptomatic patients.

For Asx versus CC, the greatest proportion of differences were found in the feature classes: GLRLM and NGTDM, whereby all features within those classes were different between asymptomatic and culprit carotid arteries. For Asx versus NC, the greatest proportion of differences were in the NGTDM feature class. There remained a greater number of differences between Asx versus CC than Asx versus NC (86.0% versus 71.0%), see Table 3.50.

Table 3.50 Multi-slice Analysis: Resegmentation - Asymptomatic versus Symptomatic

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC		
First Order (n=18)	15 (83.3%)	15 (83.3%)
GLCM (n=24)	19 (79.2%)	19 (79.2%)
GLDM (n=14)	13 (92.9%)	13 (92.9%)
GLRLM (n=16)	16 (100.0%)	16 (100.0%)
GLSZM (n=16)	12 (75.0%)	12 (75.0%)
NGTDM (n=5)	5 (100.0%)	5 (100.0%)
Total (n=93)	80 (86.0%)	80 (86.0%)
Asx versus NC		
First Order (n=18)	11 (61.1%)	9 (50.0%)
GLCM (n=24)	18 (75.0%)	17 (70.8%)
GLDM (n=14)	11 (78.6%)	11 (78.6%)
GLRLM (n=16)	12 (75.0%)	12 (75.0%)
GLSZM (n=16)	12 (75.0%)	12 (75.0%)
NGTDM (n=5)	5 (100.0%)	5 (100.0%)
Total (n=93)	69 (74.2%)	66 (71.0%)

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries; BH, Benjamini-Hochberg correction

3.3.3.3 Multi-slice analysis: asymptomatic with carotid atherosclerosis only

Table 3.51 Multi-slice analysis: Asx with plaque only

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	8 (44.4%)	5 (27.8%)
GLCM (n=24)	11 (45.8%)	9 (37.5%)
GLDM (n=14)	7 (50.0%)	5 (35.7%)
GLRLM (n=16)	5 (31.3%)	5 (31.3%)
GLSZM (n=16)	5 (31.3%)	3 (18.8%)
NGTDM (n=5)	1 (20.0%)	0 (0.0%)
Total (n=93)	37 (39.8%)	27 (29.0%)
Asx versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	5 (27.8%)	3 (16.7%)
GLCM (n=24)	9 (37.5%)	4 (16.7%)
GLDM (n=14)	6 (42.9%)	3 (21.4%)
GLRLM (n=16)	3 (18.8%)	1 (6.25%)
GLSZM (n=16)	3 (18.8%)	0 (0.05)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	26 (28.0%)	11 (11.8%)
Asx versus CC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	15 (83.3%)	15 (83.3%)
GLCM (n=24)	18 (75.0%)	18 (75.0%)
GLDM (n=14)	13 (92.9%)	13 (92.9%)
GLRLM (n=16)	14 (87.5%)	14 (87.5%)
GLSZM (n=16)	12 (75.0%)	12 (75.0%)
NGTDM (n=5)	5 (100.0%)	5 (100.0%)
Total (n=93)	77 (82.8%)	77 (82.8%)
Asx versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	9 (50.0%)	8 (44.4%)
GLCM (n=24)	16 (66.7%)	13 (54.1%)
GLDM (n=14)	10 (71.4%)	10 (71.4%)
GLRLM (n=16)	11 (68.8%)	11 (68.8%)
GLSZM (n=16)	10 (62.5%)	8 (50.0%)
NGTDM (n=5)	5 (100.0%)	5 (100.0%)
Total (n=93)	61 (65.6%)	55 (59.1%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

Table 3.51 demonstrates that when we consider asymptomatic arteries with plaque only, there were fewer differences than considering all asymptomatic carotid arteries. This agreed with the single-slice approach findings. However, in multi-slice analysis there remained substantial differences between asymptomatic carotid arteries and symptomatic carotid arteries, and there were more differences between CC vs Asx than NC vs Asx.

3.3.3.4 Multi-slice analysis: carotid plaque sub-set analysis
3.3.3.4.1 Symptomatic Patients: Culprit versus Non-culprit Carotids

In the paired analysis of carotid arteries in symptomatic patients (culprit versus non-culprit), there were 8 calcified carotid plaque type pairs, 2 noncalcified pairs and 20 mixed pairs. There were no differences between calcified pairs in multi-slice analysis nor between non-calcified pairs when using the unrestricted algorithm. There were some differences between CC vs NC in the mixed plaque types, see Table 3.52.

Following resegmentation, there was only 1 radiomic feature that had a $p<0.05$ between CC and NC: GLRLM: Long Run High Grey Level Emphasis. However, this was no longer statistically significant following Benjamini-Hochberg correction. There were no differences between noncalcified pairs, but there were between mixed pairs, see Table 3.52.

Table 3.52 Multi-slice analysis: Mixed carotid plaque type –culprit versus non-culprit

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Mixed CC versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	3 (16.7%)	1 (5.56%)
GLCM (n=24)	4 (16.7%)	2 (8.33%)
GLDM (n=14)	5 (35.7%)	3 (21.4%)
GLRLM (n=16)	4 (25.0%)	2 (12.5%)
GLSZM (n=16)	2 (12.5%)	1 (6.25%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	18 (19.4%)	9 (9.68%)
Mixed CC versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	12 (66.7%)	9 (50.0%)
GLCM (n=24)	15 (62.5%)	13 (54.2%)
GLDM (n=14)	11 (78.6%)	10 (71.4%)
GLRLM (n=16)	12 (75.0%)	8 (50.0%)
GLSZM (n=16)	6 (37.5%)	3 (18.8%)
NGTDM (n=5)	2 (40.0%)	2 (40.0%)
Total (n=93)	58 (62.4%)	45 (48.4%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

3.3.3.4.2 Inter-patient comparisons: Asymptomatic versus Symptomatic

For subset analysis according to carotid plaque type for inter-patient comparisons, there were too few noncalcified plaques (Asx: 1, CC: 2, NC: 5 carotid arteries). However, there were sufficient carotid arteries for comparisons of calcified plaques (Asx: 14, CC: 9, NC: 15) and of mixed plaques (Asx: 13, CC: 30, NC: 21 carotid arteries).

Table 3.53 Multi-slice Carotid Plaque Type Subset Analysis: Asx vs CC or NC

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Calcified Asx versus CC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	5 (27.8%)	1 (5.56%)
GLCM (n=24)	10 (41.7%)	3 (12.5%)
GLDM (n=14)	6 (42.9%)	2 (14.3%)
GLRLM (n=16)	3 (18.8%)	0 (0.0%)
GLSZM (n=16)	6 (37.5%)	2 (12.5%)
NGTDM (n=5)	2 (40.0%)	1 (20.0%)
Total (n=93)	32 (34.4%)	9 (9.68%)
Mixed Asx versus CC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	4 (22.2%)	3 (16.7%)
GLCM (n=24)	7 (29.2%)	0 (0.0%)
GLDM (n=14)	6 (42.9%)	2 (14.3%)
GLRLM (n=16)	4 (25.0%)	2 (12.5%)
GLSZM (n=16)	3 (18.8%)	1 (6.25%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	24 (25.8%)	8 (8.60%)
Calcified Asx versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	9 (50.0%)	6 (33.3%)
GLCM (n=24)	15 (62.5%)	12 (50.0%)
GLDM (n=14)	7 (50.0%)	4 (28.6%)
GLRLM (n=16)	6 (37.5%)	2 (12.5%)
GLSZM (n=16)	7 (43.8%)	7 (43.8%)
NGTDM (n=5)	3 (60.0%)	3 (60.0%)
Total (n=93)	47 (50.5%)	34 (36.6%)
Mixed Asx versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	4 (22.2%)	2 (11.1%)
GLCM (n=24)	0 (0.0%)	0 (0.0%)
GLDM (n=14)	2 (14.3%)	0 (0.0%)
GLRLM (n=16)	1 (6.25%)	0 (0.0%)
GLSZM (n=16)	2 (12.5%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	9 (9.68%)	2 (2.15%)
Calcified Asx versus CC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	1 (5.56%)	0 (0.0%)
GLCM (n=24)	5 (20.8%)	0 (0.0%)

GLDM (n=14)	2 (14.3%)	0 (0.0%)
GLRLM (n=16)	4 (25.0%)	0 (0.0%)
GLSZM (n=16)	0 (0.0%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	12 (12.9%)	0 (0.0%)
Mixed Asx versus CC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	12 (66.7%)	12 (66.7%)
GLCM (n=24)	18 (75.0%)	18 (75.0%)
GLDM (n=14)	12 (85.7%)	12 (85.7%)
GLRLM (n=16)	13 (81.3%)	13 (81.3%)
GLSZM (n=16)	10 (62.5%)	9 (56.3%)
NGTDM (n=5)	5 (100.0%)	5 (100.0%)
Total (n=93)	70 (75.3%)	69 (74.2%)
Calcified Asx versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	0 (0.0%)	0 (0.0%)
GLCM (n=24)	1 (4.17%)	0 (0.0%)
GLDM (n=14)	2 (14.3%)	0 (0.0%)
GLRLM (n=16)	3 (18.8%)	0 (0.0%)
GLSZM (n=16)	0 (0.0%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	6 (6.45%)	0 (0.0%)
Mixed Asx versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	4 (22.2%)	1 (5.56%)
GLCM (n=24)	12 (50.0%)	11 (45.8%)
GLDM (n=14)	10 (71.4%)	9 (64.3%)
GLRLM (n=16)	9 (56.3%)	8 (50.0%)
GLSZM (n=16)	10 (62.5%)	8 (50.0%)
NGTDM (n=5)	5 (100.0%)	5 (100.0%)
Total (n=93)	50 (53.8%)	42 (45.2%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

Table 3.53 demonstrates that there were differences between calcified carotid arteries and mixed carotid arteries. When using the unrestricted algorithm, there were more differences between NC vs Asx calcified carotid arteries than between CC vs Asx. Following resegmentation, there were no longer any differences between calcified carotid arteries. Resegmentation revealed greater differences between mixed plaque types. However, there were more differences between CC vs Asx mixed carotid arteries than NC vs Asx.

3.3.3.5 Multi-slice analysis: carotid stenosis severity

Table 3.54 Multi-slice analysis: carotid stenosis severity - culprit versus non-culprit

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Below 50% CC versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	0 (0.0%)	0 (0.0%)
GLCM (n=24)	0 (0.0%)	0 (0.0%)
GLDM (n=14)	0 (0.0%)	0 (0.0%)
GLRLM (n=16)	0 (0.0%)	0 (0.0%)
GLSZM (n=16)	0 (0.0%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	0 (0.0%)	0 (0.0%)
Above 50% CC versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	0 (0.0%)	0 (0.0%)
GLCM (n=24)	0 (0.0%)	0 (0.0%)
GLDM (n=14)	1 (7.14%)	0 (0.0%)
GLRLM (n=16)	0 (0.0%)	0 (0.0%)
GLSZM (n=16)	0 (0.0%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	1 (1.08%)	0 (0.0%)
Below 50% CC versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	0 (0.0%)	0 (0.0%)
GLCM (n=24)	0 (0.0%)	0 (0.0%)
GLDM (n=14)	0 (0.0%)	0 (0.0%)
GLRLM (n=16)	0 (0.0%)	0 (0.0%)
GLSZM (n=16)	0 (0.0%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	0 (0.0%)	0 (0.0%)
Above 50% CC versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	1 (5.56%)	0 (0.0%)
GLCM (n=24)	9 (37.5%)	0 (0.0%)
GLDM (n=14)	2 (14.3%)	0 (0.0%)
GLRLM (n=16)	0 (0.0%)	0 (0.0%)
GLSZM (n=16)	4 (25.0%)	0 (0.0%)
NGTDM (n=5)	2 (40.0%)	0 (0.0%)
Total (n=93)	18 (19.4%)	0 (0.0%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

In paired comparisons (CC vs NC) according to carotid stenosis using the unrestricted algorithm, there was little difference between CC and NC and no statistically significant differences following multiple comparisons correction, see Table 3.54. With resegmentation, there were 18 radiomic features with a $p<0.05$, however, these were no longer statistically significant following Benjamini-Hochberg correction.

Table 3.55 Multi-slice analysis: carotid stenosis severity: Asx vs CC or NC

Radiomic Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Below 50% Asx versus CC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	4 (22.2%)	2 (11.1%)
GLCM (n=24)	6 (25.0%)	0 (0.0%)
GLDM (n=14)	5 (35.7%)	0 (0.0%)
GLRLM (n=16)	2 (12.5%)	0 (0.0%)
GLSZM (n=16)	4 (25.0%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	21 (22.6%)	2 (2.15%)
Above 50% Asx versus CC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	11 (61.1%)	9 (50.0%)
GLCM (n=24)	14 (58.3%)	10 (41.7%)
GLDM (n=14)	10 (71.4%)	7 (50.0%)
GLRLM (n=16)	15 (93.8%)	12 (75.0%)
GLSZM (n=16)	12 (75.0%)	8 (50.0%)
NGTDM (n=5)	1 (20.0%)	1 (20.0%)
Total (n=93)	63 (67.7%)	47 (50.5%)
Below 50% Asx versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	4 (22.2%)	2 (11.1%)
GLCM (n=24)	7 (29.2%)	4 (16.7%)
GLDM (n=14)	4 (28.6%)	2 (14.3%)
GLRLM (n=16)	2 (12.5%)	0 (0.0%)
GLSZM (n=16)	3 (18.8%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	0 (0.0%)
Total (n=93)	20 (21.5%)	8 (8.60%)
Above 50% Asx versus NC: Multi-slice Analysis: Unrestricted, Bin Width = 25		
First Order (n=18)	10 (55.6%)	9 (50.0%)
GLCM (n=24)	14 (58.3%)	12 (50.0%)
GLDM (n=14)	10 (71.4%)	8 (57.1%)
GLRLM (n=16)	15 (93.8%)	13 (81.3%)
GLSZM (n=16)	11 (68.8%)	9 (56.3%)
NGTDM (n=5)	2 (40.0%)	1 (20.0%)
Total (n=93)	62 (66.7%)	52 (55.9%)
Below 50% Asx versus CC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	5 (27.8%)	5 (27.8%)
GLCM (n=24)	14 (58.3%)	13 (54.2%)
GLDM (n=14)	10 (71.4%)	10 (71.4%)
GLRLM (n=16)	11 (68.8%)	11 (68.8%)
GLSZM (n=16)	11 (68.8%)	10 (62.5%)
NGTDM (n=5)	4 (80.0%)	4 (80.0%)
Total (n=93)	55 (59.1%)	53 (57.0%)
Above 50% Asx versus CC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	8 (44.4%)	0 (0.0%)
GLCM (n=24)	1 (4.17%)	0 (0.0%)
GLDM (n=14)	6 (42.9%)	0 (0.0%)

GLRLM (n=16)	4 (25.0%)	0 (0.0%)
GLSZM (n=16)	5 (31.3%)	0 (0.0%)
NGTDM (n=5)	4 (80.0%)	0 (0.0%)
Total (n=93)	28 (30.1%)	0 (0.0%)
Below 50% Asx versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	5 (27.8%)	2 (11.1%)
GLCM (n=24)	12 (50.0%)	10 (41.7%)
GLDM (n=14)	9 (64.3%)	9 (64.3%)
GLRLM (n=16)	11 (68.8%)	9 (56.3%)
GLSZM (n=16)	8 (50.0%)	5 (31.3%)
NGTDM (n=5)	4 (80.0%)	3 (60.0%)
Total (n=93)	49 (52.7%)	38 (40.9%)
Above 50% Asx versus NC: Multi-slice Analysis: Resegmentation, Bin Width = 25		
First Order (n=18)	4 (22.2%)	0 (0.0%)
GLCM (n=24)	0 (0.0%)	0 (0.0%)
GLDM (n=14)	5 (35.7%)	0 (0.0%)
GLRLM (n=16)	3 (18.8%)	0 (0.0%)
GLSZM (n=16)	7 (43.8%)	0 (0.0%)
NGTDM (n=5)	2 (40.0%)	0 (0.0%)
Total (n=93)	21 (22.6%)	0 (0.0%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery

Table 3.55 demonstrates that there were greater differences between Asx vs CC in the above 50% category, than in the below 50% category. This was also the case for Asx vs NC when using the unrestricted algorithm. Following resegmentation, differences were lost between Asx vs CC/NC in the above 50% category but were increased for the below 50% category.

3.3.4 Summary of CT Angiography PyRadiomic Feature Differences

Table 3.56 summarises the proportion of radiomic feature differences between carotid artery types (Asx, CC and NC) in the 4 image settings investigated: (1) single-slice approach with no resegmentation (original image), (2) single-slice with resegmentation, (3) multi-slice approach with original image and (4) multi-slice approach with resegmentation.

Table 3.56 Summary of CTA PyRadiomics Feature Differences between Carotid Artery Types

Image Setting		Comparisons	Proportion with $p<0.05$	Proportion statistically significant following BH correction
SS	Original	Asx vs CC	51 (54.8%)	41 (44.1%)
		Asx vs NC	50 (53.8%)	35 (37.6%)
		CC vs NC	5 (5.4%)	0 (0.0%)
SS	Resegmentation	Asx vs CC	62 (66.7%)	60 (64.5%)
		Asx vs NC	40 (43.0%)	36 (38.7%)
		CC vs NC	22 (23.7%)	0 (0.0%)
MS	Original	Asx vs CC	50 (53.8%)	45 (48.4%)
		Asx vs NC	51 (54.8%)	42 (45.2%)
		CC vs NC	14 (15.1%)	3 (3.2%)
MS	Resegmentation	Asx vs CC	80 (86.0%)	80 (86.0%)
		Asx vs NC	69 (74.2%)	66 (71.0%)
		CC vs NC	57 (61.3%)	54 (58.1%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery; SS, single slice approach; MS, multi-slice approach.

Figure 3.41 and Figure 3.42 provide a visual summary of the PyRadiomic feature values for the individual carotid arteries, in different carotid artery types. The values of the 93 radiomic features were scaled to [0,1] for comparability and their values are displayed as a heatmap. Each row represents the radiomic profile of an individual carotid artery. Each figure has three panels with the 1st panel displaying the radiomic profile of the culprit carotid arteries (n=41), the 2nd panels demonstrates the profiles of the non-culprit carotid arteries (n=41) and the 3rd panel displays the radiomic profiles of the asymptomatic carotid arteries (n=50). Visually, we can see that there are greater differences between asymptomatic carotid arteries and symptomatic carotid arteries, than in paired comparisons of culprit versus non-culprit carotid arteries. In the heatmap, feature values that are low appear darker, whereas higher values are associated with brighter colours. We can see that applying resegmentation shifts the values of extracted radiomic features to higher values than without resegmentation.

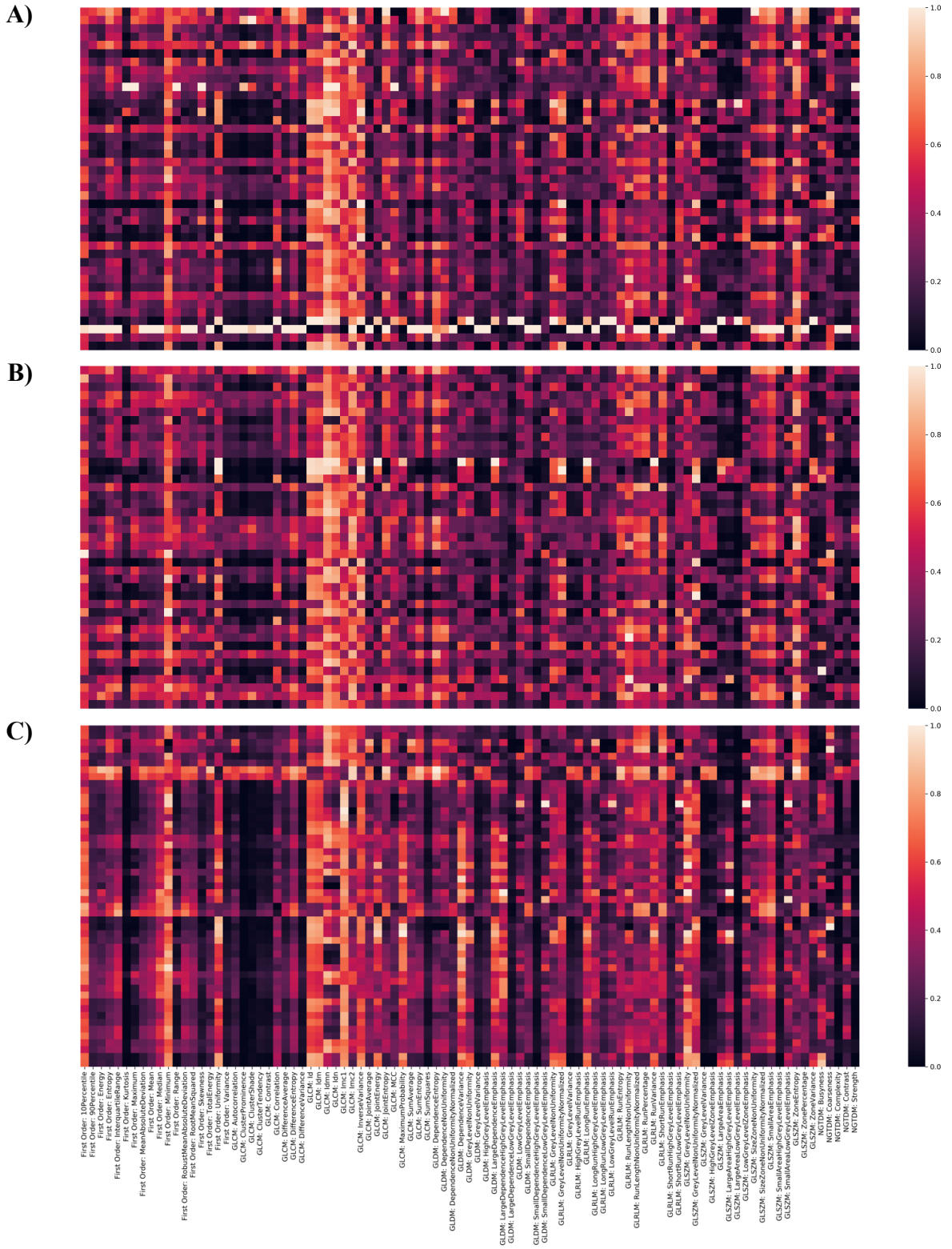


Figure 3.41 Multi-slice Analysis: Original - Heat map of Radiomic Feature Values

Heat maps of the radiomic feature values for A) culprit carotid arteries (n=41), (B) non-culprit carotid arteries (n=41) and C) asymptomatic carotid arteries (n=50) that have been normalised to a scale of 0 to 1 for comparability. In each panel (A-C), each row indicates the radiomic profile of an individual carotid artery. The brighter the colour, the higher the radiomic feature value – a colour bar is presented on the side for comparison. Each column represents the radiomic feature value for a PyRadiomics feature.

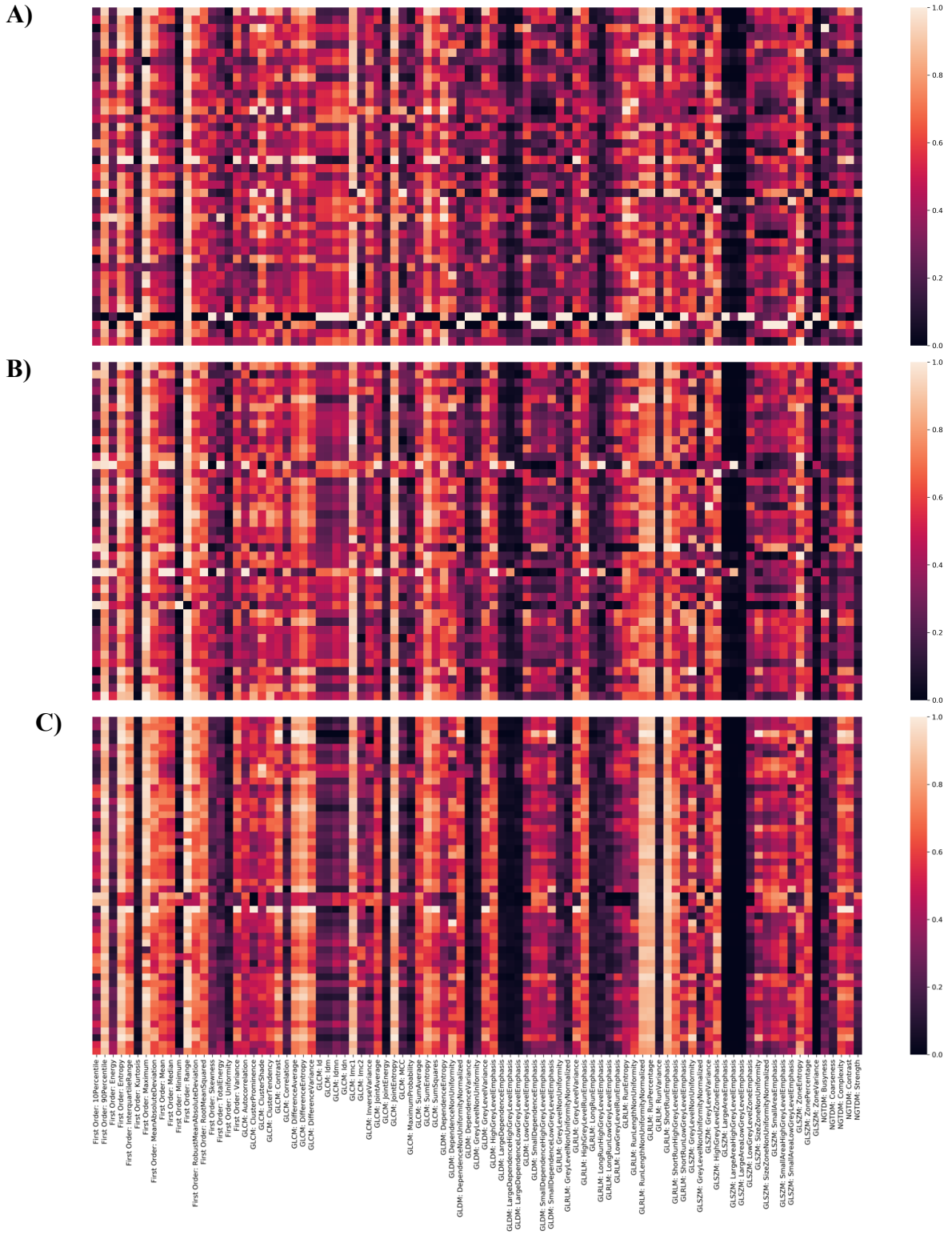
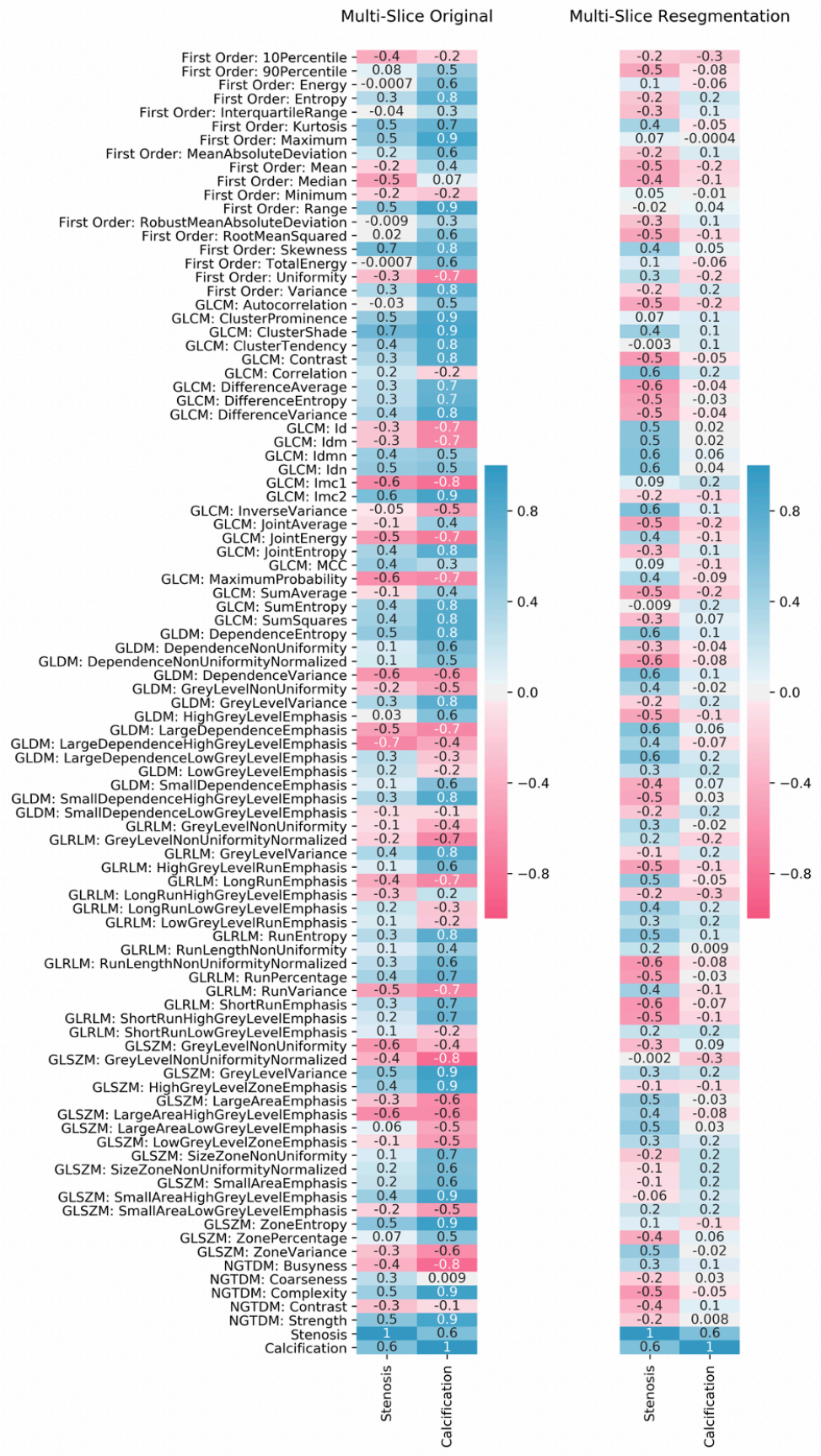


Figure 3.42 Multi-slice Analysis: Resegmentation - Heat map of Radiomic Feature Values

Heat maps of radiomic feature values for A) culprit carotid arteries (n=41), (B) non-culprit carotid arteries (n=41) and C) asymptomatic carotid arteries (n=50) extracted from volumes-of-interest derived from multi-slice analysis of carotid CT angiography scans. Resegmentation was applied to the binary masks to base radiomic feature extraction from the range of 0 to 200 Hounsfield units.

3.3.5 CTA Radiomic Feature Correlations with Carotid Stenosis and Calcium Burden

Radiomic features extracted in the multi-slice approach using PyRadiomics were not independent from carotid stenosis or carotid calcification. This is similar to the findings of Chapter 2 and the TexRAD texture features. When no resegmentation was applied, there were weak-to-moderate correlations between the radiomic features and the degree of carotid stenosis and strong correlations with carotid calcium burden. However, following resegmentation to [0, 200] HU, these correlations were attenuated. This was especially effective in minimising the correlation with carotid calcium. Figure 3.43 provides the correlations between the radiomic features, carotid stenosis and carotid calcium with and without resegmentation.



Spearman Rank Correlations were calculated between 93 radiomic features and the degree of carotid stenosis and carotid calcification for culprit, non-culprit and asymptomatic carotid arteries. The numbers in each cell represent Spearman's rho.

Figure 3.43 CTA Radiomic Feature Correlations with Carotid Stenosis and Calcification

3.3.6 Unenhanced CT

Radiomic features were extracted from unenhanced CT scans as well to assess whether higher-order radiomic features could reveal more differences than first order radiomic features which could not differentiate between culprit versus non-culprit carotid arteries with TexRAD, discussed in Chapter 2, section 2.3.8. As we have previously shown that multi-slice analysis is more effective at identifying differences between carotid artery types (see Chapter 2, section 2.3.7 and Chapter 3, section 3.3.2), we investigated the multi-slice approach for unenhanced CT with and without resegmentation.

3.3.6.1 Original Image (TexRAD unrestricted algorithm equivalent)

Without resegmentation, there were no unenhanced CT radiomic feature differences between culprit versus non-culprit carotid arteries out of the 93 PyRadiomic features calculated, even with a multi-slice approach. When comparing asymptomatic with culprit, and asymptomatic with non-culprit carotid arteries however, there were substantial differences (88.2% and 90.3% respectively), see Table 3.57. There was a slightly greater proportion of differences between Asx versus NC, than Asx versus CC with unenhanced CT – the opposite pattern to contrast-enhanced CT.

Table 3.57 Unenhanced CT - Multi-Slice Original: Asymptomatic versus Symptomatic

Feature Class	Proportion with $p<0.05$	Proportion statistically significant following BH correction
Asx versus CC		
First Order (n=18)	16 (88.9%)	16 (88.9%)
GLCM (n=24)	21 (87.5%)	21 (87.5%)
GLDM (n=14)	13 (92.9%)	13 (92.9%)
GLRLM (n=16)	14 (87.5%)	13 (81.3%)
GLSZM (n=16)	15 (93.8%)	15 (93.8%)
NGTDM (n=5)	4 (80.0%)	4 (80.0%)
Total (n=93)	83 (89.2%)	82 (88.2%)
Asx versus NC		
First Order (n=18)	16 (88.9%)	16 (88.9%)
GLCM (n=24)	21 (87.5%)	21 (87.5%)
GLDM (n=14)	13 (92.9%)	13 (92.9%)
GLRLM (n=16)	15 (93.8%)	15 (93.8%)
GLSZM (n=16)	16 (100.0%)	15 (93.8%)
NGTDM (n=5)	4 (80.0%)	4 (80.0%)
Total (n=93)	85 (91.4%)	84 (90.3%)

3.3.6.2 Resegmented Image (TexRAD restricted I algorithm equivalent)

Multi-slice analysis with resegmentation of unenhanced CT scans identified 4 radiomic features with p<0.05 when comparing culprit versus non-culprit carotid arteries, shown in Table 3.58. These were not statistically significant following Benjamini-Hochberg correction.

Table 3.58 Unenhanced CT - Multi-slice with Resegmentation: Culprit versus Non-culprit

Radiomic Feature	Culprit Carotids (n=41)	Non-culprit Carotids (n=41)	P-value
First Order: Range	199.1 (197.1, 199.5)	198.3 (196.0, 199.1)	0.033
First Order: Maximum	199.2 (197.2, 199.7)	198.7 (196.4, 199.5)	0.043
GLCM: Imc1	-0.058 (-0.075, -0.050)	-0.054 (-0.071, -0.046)	0.019
GLSZM: Zone Entropy	5.13 (5.03, 5.26)	5.09 (4.82, 5.22)	0.028

The proportion of differences between Asx and CC, as well as Asx and NC was slightly reduced following resegmentation in unenhanced CT. However, the proportions remained similar (e.g. Asx vs CC 88.2% versus 87.1% without resegmentation and with resegmentation respectively).

Table 3.59 Unenhanced CT - Multi-slice with Resegmentation: Asymptomatic versus Symptomatic

Radiomic Feature Class	Proportion with *p*<0.05	Proportion statistically significant following BH correction
Asx versus CC		
First Order (n=18)	16 (88.9%)	16 (88.9%)
GLCM (n=24)	20 (83.3%)	20 (83.3%)
GLDM (n=14)	13 (92.9%)	13 (92.9%)
GLRLM (n=16)	14 (87.5%)	14 (87.5%)
GLSZM (n=16)	16 (100.0%)	16 (100.0%)
NGTDM (n=5)	3 (60.0%)	2 (40.0%)
Total (n=93)	82 (99.2%)	81 (87.1%)
Asx versus NC		
First Order (n=18)	14 (77.8%)	14 (77.8%)
GLCM (n=24)	20 (83.3%)	20 (83.3%)
GLDM (n=14)	13 (92.9%)	13 (92.9%)
GLRLM (n=16)	15 (93.8%)	14 (87.5%)
GLSZM (n=16)	16 (100.0%)	16 (100.0%)
NGTDM (n=5)	2 (40.0%)	2 (20.0%)
Total (n=93)	80 (86.0%)	79 (84.9%)

Asx, asymptomatic; CC, culprit; NC, non-culprit; BH, Benjamini-Hochberg correction

Table 3.60 Summary of Unenhanced CT Radiomic features

Image Setting		Comparisons	Proportion with $p<0.05$	Proportion statistically significant following BH correction
MS	Original	Asx vs CC	83 (89.2%)	82 (88.2%)
		Asx vs NC	85 (91.4%)	84 (90.3%)
		CC vs NC	0 (0.0%)	0 (0.0%)
MS	Resegmentation	Asx vs CC	82 (99.2%)	81 (87.1%)
		Asx vs NC	80 (86.0%)	79 (84.9%)
		CC vs NC	4 (4.3%)	0 (0.0%)

BH, Benjamini-Hochberg correction for multiple comparisons; Asx, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery; MS, multi-slice approach.

Table 3.60 provides a summary of the proportion of radiomic features that were different between carotid artery types in the different image settings investigated for unenhanced CT, namely (1) multi-slice analysis without resegmentation (original) and (2) multi-slice analysis with resegmentation.

Figure 3.44 and Figure 3.45 provide a visual representation of the radiomic feature values derived from unenhanced CT, normalised to a range of 0 to 1 for comparability on the heat map. Each row represents the radiomic profile of an individual carotid artery. The three panels in each figure correspond to a carotid artery type in the following order: (1) culprit carotid arteries, (2) non-culprit carotid arteries and (3) asymptomatic carotid arteries. Without resegmentation in Figure 3.44, there are clear visual differences between asymptomatic carotid arteries versus the symptomatic carotid arteries. It is more difficult to identify differences visually between the panels for culprit versus non-culprit carotid arteries. With resegmentation, there is a shift in the feature values, however, CC and NC panels remain visually more similar, than when compared with the asymptomatic carotid artery panel.

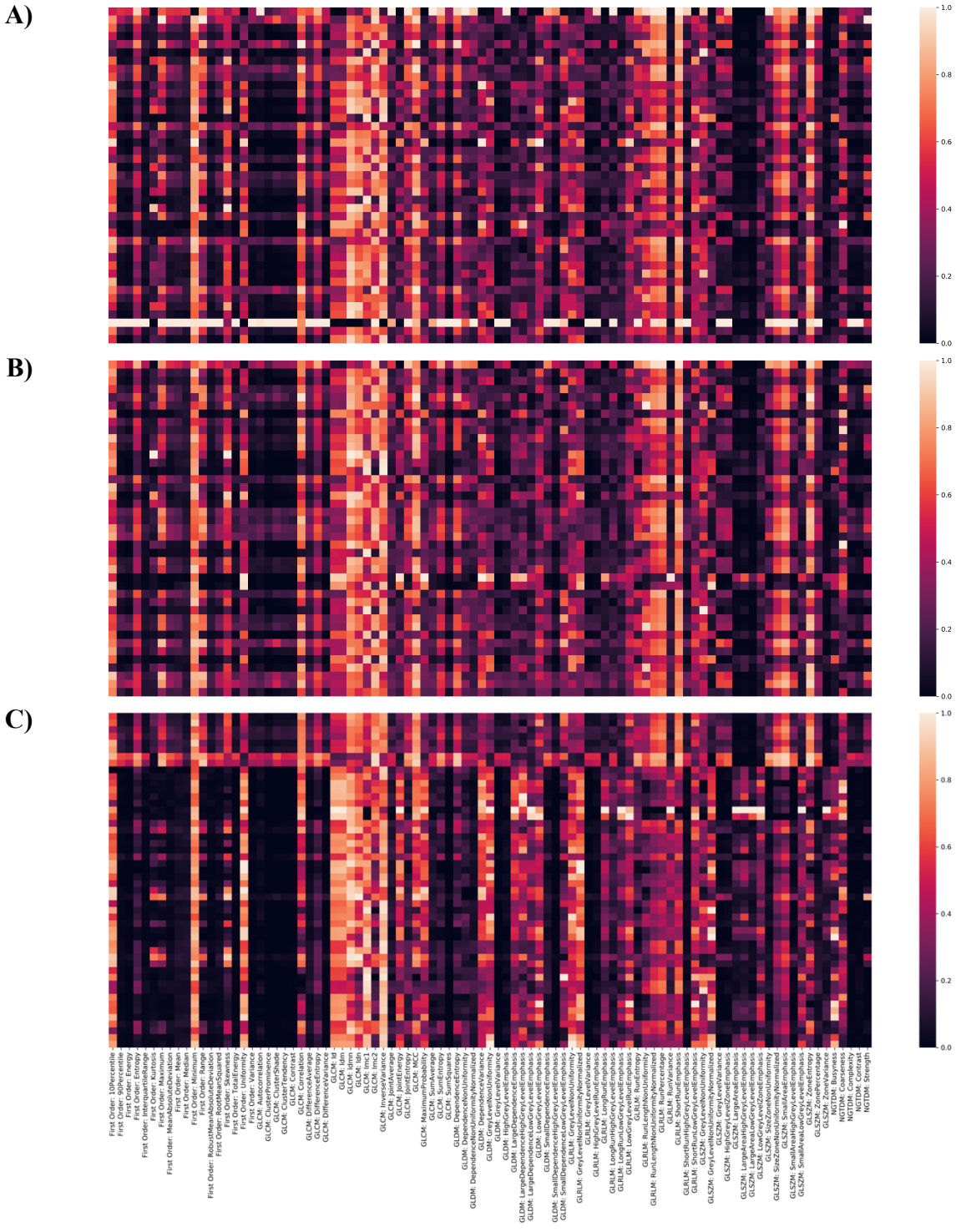


Figure 3.44 Multi-Slice Original - Heatmap of Unenhanced CT Radiomic Feature Values

Heat maps of the radiomic feature values for A) culprit carotid arteries (n=41), (B) non-culprit carotid arteries (n=41) and C) asymptomatic carotid arteries (n=50) that have been normalised to a scale of 0 to 1 for comparability. In each panel (A-C), each row indicates the radiomic profile of an individual carotid artery.

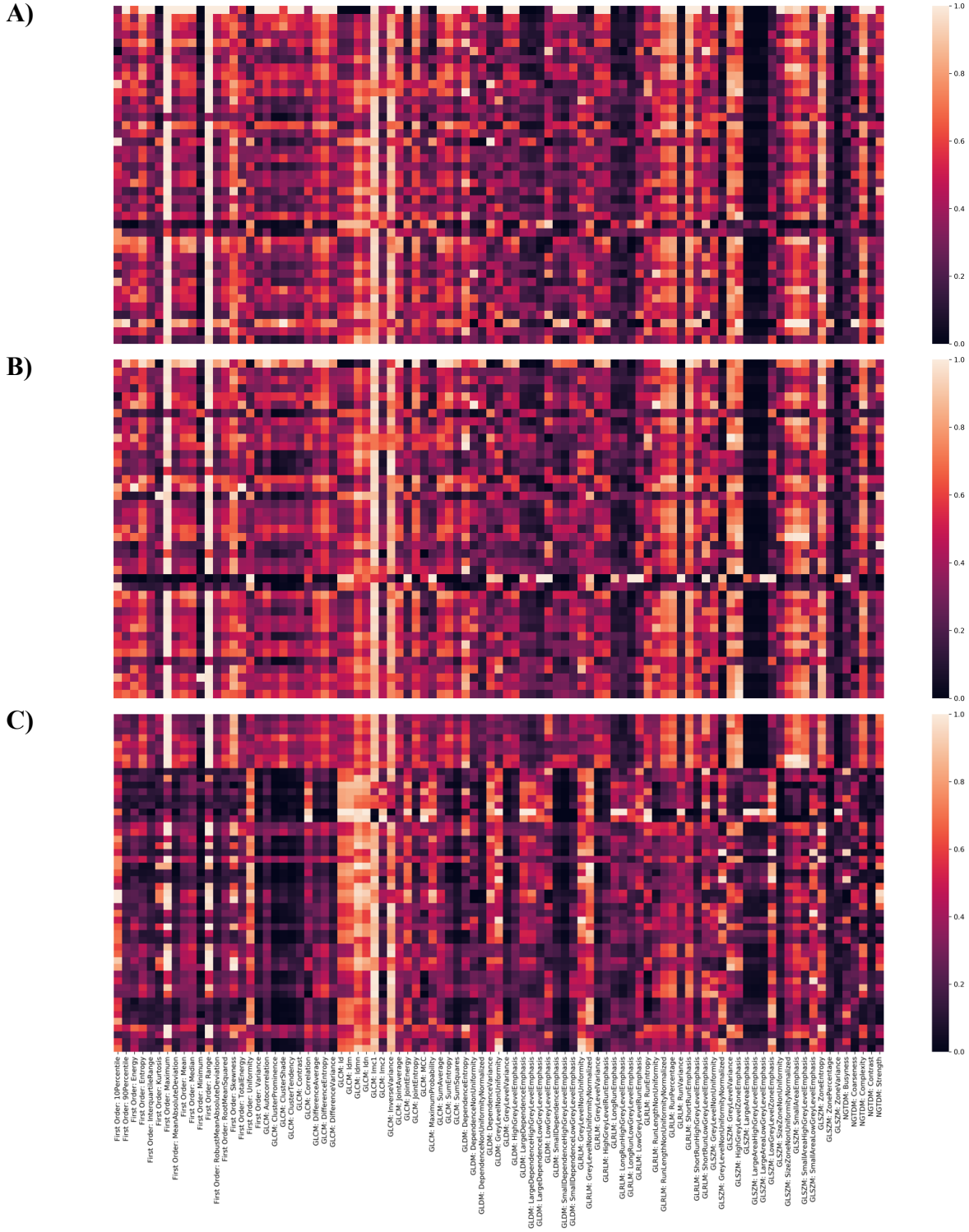


Figure 3.45 Multi-Slice: Resegmentation – Heatmap of Unenhanced CT Radiomic Features Values

Heat maps of the radiomic feature values for A) culprit carotid arteries (n=41), (B) non-culprit carotid arteries (n=41) and C) asymptomatic carotid arteries (n=50) that have been normalised to a scale of 0 to 1 for comparability. In each panel (A-C), each row indicates the radiomic profile of an individual carotid artery.

3.3.7 CT Radiomic Feature Correlations with Carotid Stenosis and Calcium Burden

Figure 3.46 Unenhanced CT Radiomic Feature Correlations with Carotid Stenosis and Calcification (Spearman Rank)

Figure 3.46 demonstrates that the radiomic features extracted from unenhanced CT scans were greatly correlated with carotid calcification, and this was not decreased following resegmentation. There were also moderate correlations with carotid stenosis with and without resegmentation.

3.3.8 Unenhanced CT versus Carotid CT Angiography: PyRadiomics

ICC values for absolute agreement and consistency were calculated to assess the concordance between PyRadiomic features derived from each imaging type. In terms of absolute agreement, there was generally poor agreement between radiomic features extracted from unenhanced CT versus those from CT angiography scans. Without resegmentation, unenhanced CT radiomic features had moderate consistency with CTA radiomic features, but this was worsened following resegmentation, see Table 3.61.

Table 3.61 Intraclass Correlation Coefficients between PyRadiomic Features from CT Angiography scans versus Unenhanced CT images

Radiomic Feature	ICC absolute agreement	ICC consistency
Original Image		
First Order: Mean	0.303 (-0.054, 0.669)	0.793 (0.696, 0.861)
GLCM: Autocorrelation	0.511 (-0.095, 0.804)	0.792 (0.695, 0.861)
GLDM: Dependence Entropy	0.309 (-0.080, 0.659)	0.712 (0.586, 0.804)
GLRLM: Grey Level NonUniformity	0.149 (-0.063, 0.435)	0.539 (0.365, 0.676)
GLSZM: Large Area Emphasis	0.027 (-0.119, 0.190)	0.038 (-0.180, 0.252)
NGTDM: Complexity	0.659 (0.070, 0.854)	0.800 (0.705, 0.866)
Resegmentation		
First Order: Mean	0.043 (-0.037, 0.166)	0.273 (0.061, 0.462)
GLCM: Autocorrelation	0.026 (-0.036, 0.110)	0.120 (-0.098, 0.328)
GLDM: Dependence Entropy	-0.057 (-0.205, 0.113)	-0.081 (-0.292, 0.137)
GLRLM: Grey Level NonUniformity	0.115 (-0.053, 0.369)	0.511 (0.332, 0.655)
GLSZM: Large Area Emphasis	0.007 (-0.158, 0.186)	0.008 (-0.208, 0.224)
NGTDM: Complexity	0.137 (-0.063, 0.338)	0.244 (0.029, 0.437)

Single measures ICC and 95% confidence interval reported; ICCs calculated using culprit and non-culprit carotid multi-slice analysis texture features

3.4 Discussion

3.4.1 Summary of key findings

In this study, we show that carotid radiomic features are quantitative features that can be derived from contrast-enhanced and unenhanced carotid CT images. We have shown differences in the radiomic profiles of culprit versus non-culprit carotid arteries, as well as differences between asymptomatic carotids versus symptomatic carotid arteries. These differences are not limited to first order texture features but extend to higher-order radiomic features derived from matrices that consider the spatial interrelationships between pixels. Using a multi-slice approach with resegmentation to analyse carotid CT angiography scans was the most effective setting for revealing differences in the radiomic features between the 3 carotid artery types: (1) asymptomatic, (2) culprit and (3) non-culprit carotid arteries. In particular, this was most effective for revealing differences between carotid arteries with mixed plaque types, and in those with below 50% carotid stenosis. However, this was not beneficial in revealing differences in calcified carotid plaque types and in those above 50% carotid stenosis.

3.4.2 Interpretation of findings

3.4.2.1 TexRAD versus PyRadiomics

There were moderate-to-good levels of agreement between the first order features calculated in TexRAD and those calculated in PyRadiomics as determined by the intraclass correlation coefficient. The discrepancies were due to systematic differences in rounding, image quantisation and/or feature definitions used, for example TexRAD uses the natural logarithm to calculate ‘entropy’, whereas PyRadiomics uses log base 2. This demonstrates that radiomic features extracted by different radiomic packages, even with the same name, must be interpreted with caution, since they may not be directly comparable. Nevertheless, since the differences between TexRAD and PyRadiomics for first order features were systematic, the statistics and correlations observed between culprit, non-culprit and asymptomatic carotid arteries were similar.

Each radiomic software package has its advantages and disadvantages. Since TexRAD is a commercial research software, it has been designed to optimise the user experience and has a simple-to-use interface that allows for manual ROI segmentation, texture feature extraction

and data storage all in one environment. TexRAD is easier for clinicians to use immediately (compared with PyRadiomics) and even though the TexRAD texture features that are calculated are more limited in number than other radiomics solutions, they have been widely investigated in the literature[67,77,163]. However, the radiomic features calculated in TexRAD are first order features which consider the grey level intensities of the ROI, rather than spatial information between neighbouring pixels. PyRadiomics enables the extraction of first order features, as well as those derived from matrices that consider spatial relationships and therefore more information can be derived from the ROI. PyRadiomics is not as intuitive to use for a clinician, however it is available as a plugin with the open-source software 3D Slicer to accommodate for this.

3.4.2.2 CT angiography Radiomics and Effect of Resegmentation

In the single-slice approach, without resegmentation, there were 5 radiomic features with a p-value<0.05 when comparing culprit versus non-culprit carotid arteries. Of these 5 radiomic features, only 1 was from the first order feature class (First Order: Median), the others were higher-order features. In contrast, with TexRAD, using a single-slice approach without resegmentation (i.e. when using the unrestricted algorithm), there were no differences between the texture features of culprit versus non-culprit carotid arteries. In this image setting, however, these 5 radiomic features were no longer statistically significantly different after correction for multiple comparisons.

Following resegmentation, which restricts the range of Hounsfield units considered within the ROI for radiomic feature calculation, the number of differences between culprit and non-culprit carotid arteries increased. The benefit of resegmentation with PyRadiomics echoed that seen with TexRAD and texture analysis of carotid CT angiography scans, (Chapter 2, Section 2.3). In this chapter, resegmentation limited the range to [0, 200 HU] which was equivalent to the restricted I algorithm investigated in TexRAD.

Complimenting the findings of Chapter 2, multi-slice analysis improved the ability to detect differences between culprit versus non-culprit carotid arteries, even without resegmentation, compared with the single-slice approach. With resegmentation, 58.1% of the extracted radiomic features were statistically significantly different in CC versus NC comparisons. The PyRadiomics first order culprit carotid arteries findings were similar to those of TexRAD: (1)

culprit carotid arteries have a lower First Order: Mean versus non-culprit carotid arteries, (2) culprit carotid arteries have a higher First Order: Kurtosis and (3) and culprit carotids have a higher First Order: Skewness. Whilst First Order: Variance (equivalent to TexRAD's 'standard deviation' squared and which was higher in non-culprit carotid arteries) was not statistically significantly different with PyRadiomics and multi-slice analysis with resegmentation, First Order: Interquartile Range was higher in non-culprit versus culprit carotid arteries.

Multi-slice analysis with resegmentation was also the optimal setting for detecting differences between the carotid arteries of asymptomatic patients and symptomatic patients. There were more differences between asymptomatic carotid arteries and culprit carotid arteries (86.0% of features), versus between asymptomatic and non-culprit carotid arteries (71.0%).

In this study, a lower limit of 0 HU was used, and an upper limit of 200 HU was used for resegmentation, mirroring the specifications for the restricted I algorithm used with TexRAD analysis in Chapter 2. This resegmentation range for CT angiography analysis, decreased the effect of carotid calcification and luminal contrast on the extracted texture features as evidenced by their decrease in correlations with PyRadiomics features following resegmentation. We originally hypothesised that resegmentation would be more important for first order features which are calculated based on the histogram of grey level intensities. However, there were moderate-to-high correlations between carotid calcification and radiomic features from several feature classes, including GLCM and NGTDM without resegmentation. All radiomic features derived from CTA scans were found to benefit from resegmentation.

The radiomic features extracted following resegmentation [0, 200] HU describes patterns found with the positive pixels within the image, up to 200 Hounsfield units. Accordingly, within this range of Hounsfield units we may be capturing information relating to, but not limited to, different plaque components. Different attenuation values have been used to identify various components of atherosclerotic carotid plaque on CT angiography images by de Weert et al., whereby calcification: >130 HU, fibrous tissue: 60-130 HU and lipid core: <60 HU[179].

3.4.2.3 Measuring 'Heterogeneity'

It is difficult to evaluate whether culprit or non-culprit carotid arteries are associated with greater heterogeneity. The radiomic features extracted from the different feature classes measure variation in several aspects, some are higher in culprit carotid arteries, whilst others are higher in non-culprit carotid arteries.

Without resegmentation, 14 radiomic features had a p-value<0.05 between culprit versus non-culprit carotid arteries in multi-slice analysis, although only 3 remained statistically significant following correction for multiple comparisons: (1) culprit carotid arteries had a lower First Order: Median than non-culprit carotid arteries, (2) culprit carotid arteries had a lower GLDM: Large Dependence High Grey Level Emphasis and (3) culprit carotid arteries had a lower GLSZM: Large Area High Grey Level Emphasis than non-culprit carotid arteries. The lower First Order: Median in culprit carotid arteries, likely reflects how culprit carotid arteries in this study were more likely to have a higher degree of carotid stenosis. The texture feature had a moderate negative correlation with carotid stenosis (rs=-0.5) and it was not correlated with the degree of carotid calcification (rs=0.07). GLDM: Large Dependence High Grey Level Emphasis involves measuring the large dependencies between high grey level values, with a greater value associated with more homogeneous textures amongst high grey level intensities. This texture feature was lower in culprit carotid arteries versus non-culprit carotid arteries suggesting that there was less homogeneity in high grey level intensities for culprit carotid arteries. It was negatively correlated with carotid stenosis (rs=-0.7) and with carotid calcium burden (rs=-0.4). Finally, GLSZM: Large Area High Grey Level Emphasis which quantifies the distribution of zones (i.e. connected voxels) with the high grey level intensities, was lower in culprit carotid arteries and was negatively correlated with the degree of carotid stenosis and calcification (rs=-0.6). Overall, without resegmentation, the radiomic features are closely related to carotid stenosis and/or calcification and there is little indication in terms of heterogeneity. Please note that the presence of a higher degree of carotid stenosis in the culprit carotid artery is unsurprising in this study due to the nature of the inclusion criteria for the previous vascular imaging studies from which this pooled dataset was derived, for example, the ICARUSS study recruited patients with ≥50% carotid stenosis in the culprit carotid artery[151].

With resegmentation, as implemented in this study, we extracted information from the images that were less dependent on carotid stenosis and calcification. GLCM radiomic features are calculated from the grey level co-occurrence matrix which counts voxel pairs with certain grey level values in specific directions and distances[82]. In multi-slice analysis with resegmentation, culprit carotid arteries were associated with a higher GLCM: Cluster Shade which implies greater asymmetry about the mean. However, non-culprit carotid arteries were associated with higher GLCM: Difference Entropy, a measure of the variability in neighbourhood grey level intensity value differences. In addition, culprit carotid arteries had a higher GLCM: Inverse Difference Moment, a measure of local homogeneity whereby a higher value indicates greater uniformity in the local grey level values[180].

The GLDM measures the number of connected voxels dependent on the centre voxel. In culprit carotid arteries, GLDM: Dependence Entropy was higher indicating a higher variability of grey level dependencies within the ROIs than in non-culprit carotid arteries. GLDM: Grey Level NonUniformity was also higher in culprit carotid arteries suggesting there was decreased similarity in grey level intensity values within the 0-200 HU range for culprit versus non-culprit carotids. However, GLDM: Small Dependence Emphasis which is associated with less homogeneous textures that higher the value was found to be lower in culprit carotid arteries than in non-culprit carotids.

Similarly, for the other second-order and higher-order feature classes, no consensus could be reached regarding the heterogeneity of the different carotid artery types. For example, culprit carotid arteries were associated with a higher GLRLM: Run Entropy which quantifies the randomness in the distribution of grey level run lengths and higher values are associated with increased heterogeneity in texture patterns. Meanwhile, culprit carotid arteries were associated with a lower NGTDM: Complexity than non-culprit carotid arteries. A NGTDM is based on the differences between each voxel and its neighbouring voxels and a higher NGTDM: Complexity is associated with an image that is non-uniform.

3.4.2.4 Subset Analyses: Asymptomatic with carotid atherosclerosis only

In both single-slice and multi-slice approaches, when considering asymptomatic carotid arteries with plaque only (compared to using all asymptomatic carotid arteries), there was a decrease in the number of differences between Asx vs CC or NC carotid arteries. However, there remained to be differences when using the unrestricted algorithm and following resegmentation. This suggests that certain radiomic features are reflecting characteristics of the carotid plaque, but also that there are radiomic features that reflect characteristics of the carotid artery beyond the plaque itself.

3.4.2.5 Subset Analyses: Carotid Plaque and Carotid Stenosis Severity

When stratifying by carotid plaque type in the intra-patient comparisons of CC vs NC carotid arteries, there were no differences between paired calcified carotid plaques. This is consistent with the findings of Chapter 2 and TexRAD features. However, there were differences between paired mixed carotid plaque types, with more differences being revealed following resegmentation. For inter-patient comparisons of Asx vs CC or NC, there were too few noncalcified plaques to carry out a subset analysis, but sufficient for calcified and mixed plaques. There were differences between Asx vs CC/NC calcified carotid arteries when using the unrestricted algorithm, but these differences were lost following resegmentation. This is concordant with how resegmentation limits the effect of calcium on the extracted radiomic features. There were consistently differences between mixed plaque types with and without resegmentation. This suggests that radiomic analysis is potentially most useful when dealing with mixed carotid plaques.

When stratifying carotid arteries according to carotid stenosis severity, there were few differences between CC vs NC in the below 50% category. In the above 50% category, there were 18 radiomic features in multi-slice analysis that had a $p<0.05$ between CC and NC, but this was no longer statistically significant following Benjamini-Hochberg correction. However, there were differences in below 50% and above 50% categories between Asx vs CC and Asx vs NC. This suggests that radiomic features go beyond reflecting luminal stenosis and that culprit and non-culprit carotid arteries have more similar radiomic profiles than asymptomatic carotid arteries and symptomatic carotid arteries. However, there were only 3 asymptomatic carotid arteries with stenosis above 50% and so further interpretation is limited.

3.4.2.6 Unenhanced CT

No differences were found between culprit versus non-culprit carotid arteries in multi-slice analysis without resegmentation. 4 radiomic features had $p<0.05$ following resegmentation: (1) culprit carotid arteries had a larger First Order: Range than non-culprit carotid arteries, (2) culprit carotids had a higher First Order: Maximum, (3) culprit carotids had a lower GLCM: Imc1 and (4) a higher GLSZM: Zone Entropy. For GLCM: Imc1 (informational measure of correlation), the closer the values to 0, the less mutual information there is, and this is associated with higher complexity. However, a higher value GLSZM: Zone Entropy suggests there is greater heterogeneity in texture patterns. However, these 4 radiomic features were no longer statistically different following correction for multiple comparisons.

Radiomic features extracted from unenhanced CT scans were moderately to highly correlated with carotid stenosis and carotid calcification. Resegmentation in this case was not sufficient to decrease the correlation between radiomic features and stenosis and/or calcification, unlike for CTA radiomic features. Resegmentation actually decreased the level of agreement (both absolute and in terms of consistency) between radiomic features extracted from CTA scans and unenhanced CT scans. Overall, there was poor absolute agreement between radiomic features derived from CTA versus unenhanced CT scans, and there was a range from poor-to-good consistency as determined by ICC measures, depending on the radiomic feature assessed. This demonstrates that radiomic features are specific to imaging modality and whether or not contrast was used.

Due to the high correlation of unenhanced CT radiomic features with carotid calcification, it is likely the differences identified between asymptomatic versus culprit and asymptomatic versus non-culprit carotid arteries reflected the differences in carotid calcification between them. Asymptomatic carotid arteries were associated with a lower carotid calcium score than symptomatic carotid arteries. Meanwhile there were no differences in carotid calcification between culprit versus non-culprit carotid arteries.

3.4.3 Limitations

The multiplicity of analyses is a limitation as this increases the likelihood of false positive results. This has been controlled for with the Benjamini-Hochberg correction for multiple comparisons and is less conservative than the Bonferroni correction to reduce the risk of a Type I error. In addition, we have only investigated unfiltered radiomics features in this study. There are other radiomic features that could be extracted following image filtering, Gabor filters or wavelet transformations. However, a review article found that most radiomic studies have extracted radiomic features from CT imaging without any pre-processing or normalisation[169]. Furthermore, as this was a proof-of-principle study using first order and higher-order texture features, we wanted to limit the number of features extracted to enable a better in-depth analysis of texture feature differences. Future work will investigate the extraction of further radiomic features following image filtering. In addition, the radiomic features extracted in this study is not limited to carotid plaque only due to the nature of the image segmentation, we may also be capturing information about the carotid vessel tissue itself. In the future, with the development of improved carotid plaque segmentation methods, we may be able to focus on carotid plaque analysis only.

3.5 Conclusions

In this study, we shifted from TexRAD texture analysis (which focused on first order features only) to embrace the 'omics' of radiomics and studied the utility of 93 radiomic features from 6 different feature classes. Our results indicate that radiomics analysis is feasible for carotid CTA studies and that higher-order features contain useful information in the differentiation of carotid artery types. There were CTA radiomic feature differences between culprit and non-culprit carotid arteries within a symptomatic patient, and even greater differences between the carotid arteries of asymptomatic versus symptomatic patients.

Chapter 4: CT Angiography Radiomics Robustness Analysis

Chapter summary:

This chapter focuses on assessing the robustness of radiomic features to variations in the radiomics workflow, with a particular focus on variabilities within image segmentation. Computer vision morphological operations were applied to regions-of-interest (ROIs) drawn around the carotid artery to mimic human segmentation under- and over-estimations with manual delineations. This was conducted under several image settings to include resegmentation, normalisation, different quantisation methods and/or resampling methods. We sought to identify the optimal image settings and the most robust and non-redundant radiomic features for the identification of culprit versus non-culprit carotid arteries in symptomatic patients.

4.1 Overview

The radiomics workflow is a multi-step process whereby each step constitutes a possible source of variation. For example, this may be in image acquisition (e.g. use of different scanners), image pre-processing (e.g. differences in resampling methods, normalisation, resegmentation), image segmentation (e.g. intra- and inter-observer variability), radiomic feature extraction (e.g. definitions of features used) and/or model building (e.g. feature selection methods employed or ML algorithms evaluated).

Robustness analyses evaluates the impact of these variations on radiomic features, aiming to identify those most immune to such perturbations. According to biomarker discovery guidelines, a radiomic biomarker should be reproducible, robust and accurate[66] prior to use in clinical decision-making. A robust radiomic feature remains stable between two measurements when certain conditions change[81] and evaluating the robustness of radiomic features can be used as a feature selection method prior to further analysis[181] e.g. incorporation into machine learning models.

Zwanenburg identified a series of potential data analysis pitfalls that could impact the reproducibility and generalisability of radiomics models and studies[182]. Generalisability refers to models that can predict well on new datasets which are similar to, but are independent to,

the dataset from which the model was developed[182]. One of the pitfalls highlighted, was the inclusion of non-robust biomarkers which could lead to models with poor generalisability.

The majority of robustness and repeatability studies thus far have been conducted within the field of oncology, such as in non-small cell lung cancer[183] or oesophageal cancer[184]. In cardiovascular imaging, there have been relatively few studies: one used an imaging phantom in single photon emission computed tomography[181] and the other identified robust myocardial radiomic features from cardiac magnetic resonance images[185].

Since feature robustness is specific to the disease phenotype being studied and to the imaging modality used[186], there is an unmet need for the assessment of radiomic robustness in carotid CTA imaging.

4.1.1 Aims and Objectives

The specific objectives for this chapter are as follows:

Objectives:

- To assess the robustness of radiomic features to image segmentation perturbations
- To assess the impact of changing image pre-processing on the robustness of radiomic features
- To determine the optimal image settings for carotid CT angiography radiomic studies
- To identify a set of highly robust and non-redundant radiomic features for culprit versus non-culprit carotid artery identification

4.1.2 Hypotheses

- Some radiomic features are robust, whilst others are non-robust to perturbations
- Prior image normalisation is not necessary for carotid CTA radiomics

4.2 Materials and Methods

4.2.1 Carotid Datasets and Image Analysis

For robustness analysis, PyRadiomics features were extracted from the ROIs (for single-slice analysis) and VOIs (for multi-slice analysis) delineated by EPVL using the carotid CT angiography scans of 41 symptomatic patients (41 culprit and 41 non-culprit carotid arteries). Please refer to Chapter 2, section 2.2.1 for details regarding carotid CT angiography data acquisition. The methods for the manual delineation of the ROIs, generation of the VOIs and PyRadiomics texture feature extraction are detailed in Chapter 2, section 2.2.4 and Chapter 3, section 3.2.2.

4.2.2 Robustness Analysis Overview

We first investigated the robustness of 93 radiomic features (the individual radiomic features are shown in Chapter 3, Table 3.2) following ROI/VOI perturbations under different image settings to include normalisation, resegmentation, different image quantisation methods and resampling methods. We then determined 1) the optimal image settings (i.e. the settings that provided the highest proportion of radiomic features with excellent robustness) and 2) the most robust and non-redundant (i.e. not highly correlated) radiomic features for machine learning classification of culprit versus non-culprit carotid arteries. For each stage of the investigation, we tested both a single-slice and multi-slice approach as performed in the previous chapters. Figure 4.47 provides an overview of the robustness analysis workflow in this chapter.

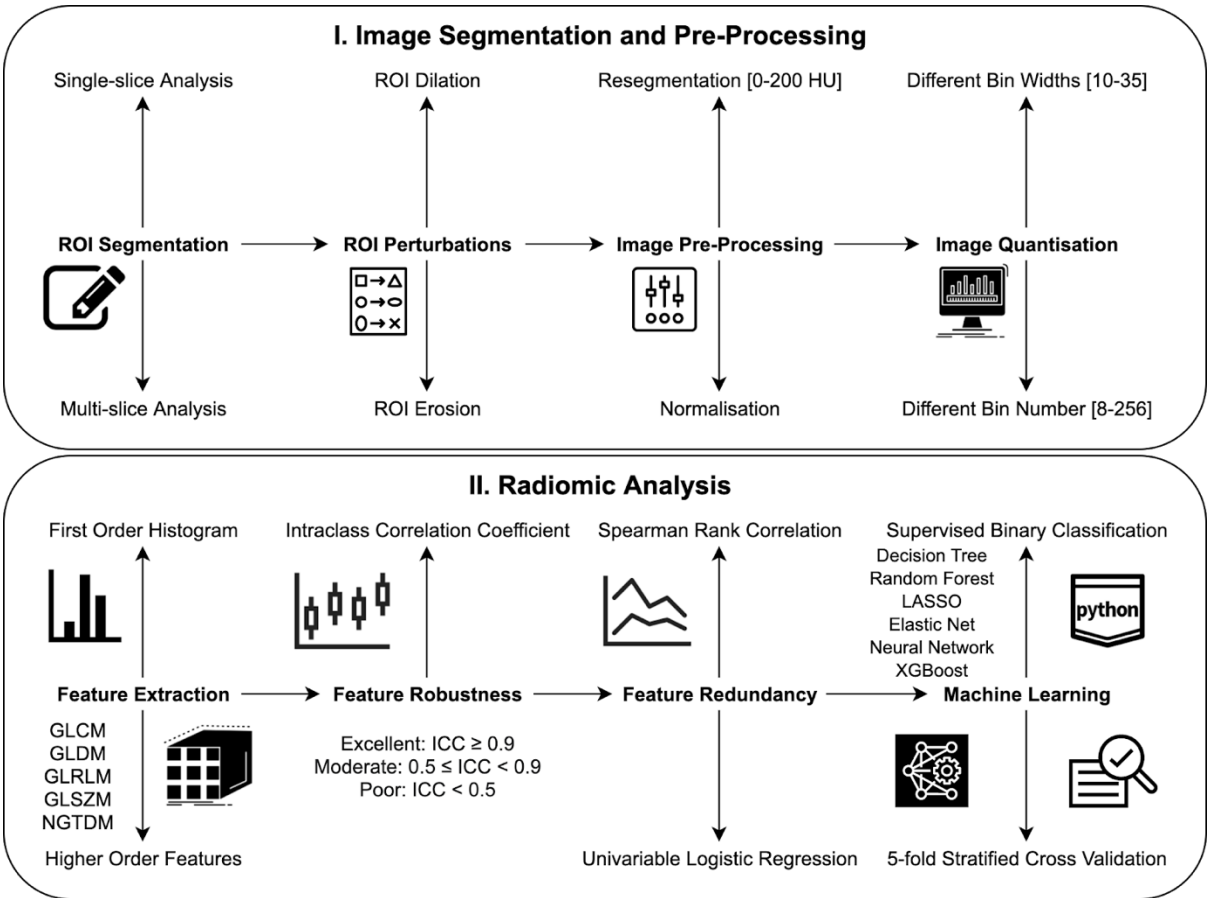


Figure 4.47 Robustness Analysis Workflow

The carotid CTA images were manually segmented to create ROIs for single-slice analysis and VOIs for multi-slice analysis. The segmented masks were subjected to morphological perturbations: dilation and erosion. The CTA images were subjected to different image settings: a) original – no image pre-processing, b) prior normalisation and c) grey value range resegmentation to [0, 200] Hounsfield units. Image quantisation refers to how the original distribution of pixel grey values in the image are rebinned. The robustness of the radiomic features were quantified using the intraclass correlation coefficient. Spearman Rank correlations were calculated to identify feature redundancy. Non-redundant radiomic features with excellent robustness were used in supervised machine learning for the differentiation of culprit versus non-culprit carotid arteries. The predictive ability of the classifiers was assessed via five-fold stratified cross-validation.

4.2.3 Segmentation Perturbations

Automatic segmentation methods are not currently widely available in many fields of medicine although this is an active area of development. Manual segmentation is therefore a key step in the current radiomics workflow. However, manual segmentation is a source of intra- and inter-observer variability since determining the ROI boundary is a subjective process, which may affect the values of extracted radiomic features.

In this study, we evaluated the impact of variations in ROI and VOI segmentations on the extracted radiomic features in a systematic manner: by performing ROI/VOI dilation and erosion with mathematical morphological operations (namely, dilation and erosion), see Figure 4.48. This was performed to simulate certain variations in ROI/VOI placement that can occur in clinical practice with human raters, such as over-estimation (simulated with the morphological operation: dilation), and under-estimation (simulated with the morphological operation: erosion).

To achieve these perturbations, the original ROIs delineated by the primary reader (EPVL) (drawn in Chapter 2, section 2.2.4), were subjected to the dilation and erosion image morphological operations implemented in Python 3.7.4, using the Python package, SciPy morphology, and its 'binary_dilation' and 'binary_erosion' functions. These functions work on the binary masks that ROIs and VOIs were converted into prior to PyRadiomic feature extraction (see Chapter 3, section 3.2.2.4).

For binary dilation, this morphological image operator gradually enlarges the boundaries of the regions of foreground pixels (i.e. those with a value of 1). This binary dilation function takes two inputs: (1) the binary mask to be dilated and (2) the structuring element (also known as a kernel) which determines the effect of the dilation on the binary mask. The larger the structuring element (determined by the radius specified, or the number of iterations [no. of repeated dilations applied]), the larger the dilation effect produced on the binary mask. Structuring elements exist in a range of shapes, including square, diamond and circular structuring elements.

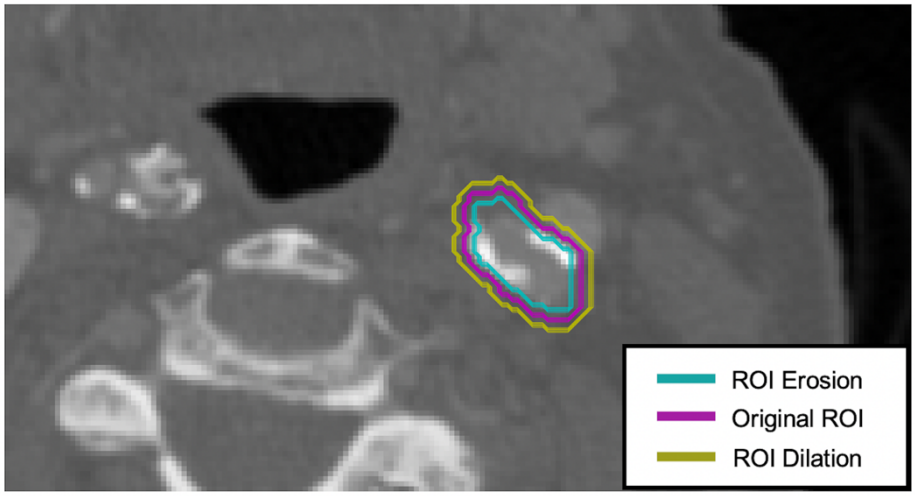


Figure 4.48 ROI segmentation and perturbations

Carotid CTA images were manually segmented to delineate the carotid artery. The original ROI was subjected to morphological operations: erosions and dilations aimed to assess the robustness of radiomic features to variations in image segmentation.

For single-slice analysis, a circular structuring element of radius 1, with iterations of 1 and 2 for ROI dilation and erosion was used, see Table 4.62. For multi-slice analysis, a spherical structuring element of radius 1, with iterations 1 and 2 for ROI dilation was used, however, only 1 iteration for ROI erosion was used in order to ensure that a sufficient number of pixels would be available for downstream radiomic feature extraction after erosion has been applied.

Table 4.62 Impact of erosion and dilation on ROI binary mask

Operation	**Erosion**		**Original**	**Dilation**	
Iterations	2	1	0	1	2
Voxel no. in ROI	244	302	364	430	500
Change in voxel no.	-120	-62	0	+66	+136
Binary mask					

Number of voxels in the ROI after erosion and dilation using circular structuring element with radius of 1; iterations n=1 and n=2. ROI, region-of-interest; no., number. Change in voxel number is in relation to original ROI.

4.2.2.2 Intra- and inter-observer image segmentation reproducibility

To assess intra-observer reproducibility in ROI delineation, 10% of the imaging dataset (n=8 carotid arteries) was selected at random and analysed on a separate occasion, 17 weeks after the initial analysis by the primary reader (EPVL). Inter-observer reproducibility was assessed as a second reader (Dr. Chris Wall, Academic Cardiology Registrar, University of Cambridge) also manually segmented the carotid arteries of 10% of the carotid CTA scans whilst blinded to the clinical data. The level of agreement between these ROI segmentations was quantified using the Dice coefficient (DC). The DC is a statistical tool that is commonly used to assess the quality of image segmentations[187]. The DC measures the level of agreement between different image segmentations by considering the level of overlap between ROI *X* and ROI *Y* over the total number of pixels in ROI *X* and ROI *Y*, according to Equation 4.5.

Equation 4.5 The Dice Coefficient

$$DC = \frac{2|X \cap Y|}{|X| + |Y|}$$

where $|\cdot|$ *denotes the cardinality of the pixels contained in a certain set.*

The DC was subsequently calculated for: (1) intra-observer reproducibility by comparing the repeated ROIs drawn by the primary reader with the original ROIs, (2) inter-observer reproducibility by comparing the primary reader's ROIs with the second reader's ROIs and (3) ROIs generated following morphological operations versus the original ROIs. The original ROIs refer to the ROIs drawn by the primary reader in the first sitting and were taken as ground truth. The distributions of the Dice coefficient values for intra- and inter-observer variability were compared with those for the ROIs generated following the perturbations. This was to assess whether the ROI variations achieved by the perturbations reflected the variability that occurred in practice with human operators and therefore be clinically relevant. The DC ranges from 0 (indicates no spatial overlap between segmentations) to 1 (indicates complete overlap)[187].

4.2.4 Image Pre-processing

As the field of radiomics in carotid CTA evaluation is nascent, there is no established best practice for specific image pre-processing schemes that should be used in carotid CTA radiomic studies. We have therefore investigated several schemes to evaluate the robustness of radiomic features to segmentation perturbations under different image pre-processing configurations. Table 4.63 outlines the different combinations of image pre-processing procedures investigated in this study.

Table 4.63 Different PyRadiomics Analysis Schemes

Analysis Scheme	Normalisation	Resegmentation
Single-slice A	-	-
Single-slice B	+	-
Single-slice C	-	+
Multi-slice A	-	-
Multi-slice B	+	-
Multi-slice C	-	+

Three schemes were investigated in single-slice analysis and multi-slice analysis respectively: a) Original image (no image pre-processing), b) Normalisation and c) Resegmentation. + indicates the respective image pre-processing operation was applied, - indicates it was not applied.

4.2.4.1 Normalisation

Prior normalisation before radiomic feature extraction changes the range of pixel intensity values. In PyRadiomics, the mean of the image is subtracted, and the pixel values are divided by the standard deviation such that the pixel values assume a Gaussian distribution with a zero-mean and unit standard deviation. Normalisation as a pre-processing step is important for MRI scans where pixel values are arbitrary, however this is not the case for CT images where the pixel values are calibrated to Hounsfield units. Nevertheless, CTA images may have differences in contrast filling which could potentially impact the robustness of radiomic features. Without prior consensus in the literature as to best practice for carotid CTA radiomics, we investigated schemes with and without prior normalisation. The CTA images were normalised by using the PyRadiomics 'normalizeImage' function.

4.2.4.2 Resegmentation

Resegmentation refers to the process whereby only pixels within a specified range are used for radiomic feature calculation within the ROI/VOI. For robustness analysis, resegmentation was applied with an upper limit of 200, and a lower limit of 0 which restricted radiomic feature extraction to only the pixels with HU between 0 and 200. This matches the restricted I algorithm in TexRAD (see Chapter 2, section 2.2.4.1) that was shown to be useful in detecting differences between carotid artery types (culprit, non-culprit or asymptomatic carotid arteries). This resegmentation setting helps to exclude the effects of excess carotid macro-calcification and limits the effect of luminal contrast and perivascular carotid fat within the CTA ROI/VOI. Please note that ROI erosion was not performed when resegmentation was used, only ROI dilation was performed in order to ensure that all ROIs had sufficient pixels for radiomic feature extraction. Similarly, normalisation and resegmentation were not applied to the image at the same time, as this would result in insufficient pixels/voxels for radiomic feature extraction when using a fixed bin width.

4.2.5 Image Quantisation

Before radiomic features are extracted from medical images, the image must be quantised, either by using a fixed number of bins (BN) or by using a fixed bin width (BW; also referred to as bin size). Image quantisation was previously described in Chapter 3, section 3.2.2.6. We varied the BWs of the image grey level histogram from 10 to 35, in increments of 5. For BN variation, we varied the fixed number of bins as follows, 8, 16, 32, 64, 128 and 256. This range of bin sizes was chosen based on the guidance in the PyRadiomics documentation for good reproducibility and performance[188]. Please note that when using the prior normalisation image setting, it was only possible to vary the number of bins rather than using different bin widths (as this led to too few grey values for radiomic feature calculation). When using the resegmentation image setting, we initially used a fixed bin width of 25 only (PyRadiomics version 3.0 default) and Table 4.64 provides a summary of the different image settings investigated. An additional analysis was later carried out in the multi-slice approach to assess the use of smaller bin widths in the resegmentation image setting (fixed bin width of 10, 15 and 20) to see if this increased the proportion of robust radiomic features.

Table 4.64 Summary of image settings tested

#	Image Type	Image Quantisation Method	Quantisation Value
1	Original	Bin Number	8
2	Original	Bin Number	16
3	Original	Bin Number	32
4	Original	Bin Number	64
5	Original	Bin Number	128
6	Original	Bin Number	256
7	Original	Bin Width	10
8	Original	Bin Width	15
9	Original	Bin Width	20
10	Original	Bin Width	25
11	Original	Bin Width	30
12	Original	Bin Width	35
13	Prior Normalisation	Bin Number	8
14	Prior Normalisation	Bin Number	16
15	Prior Normalisation	Bin Number	32
16	Prior Normalisation	Bin Number	64
17	Prior Normalisation	Bin Number	128
18	Prior Normalisation	Bin Number	256
19	Resegmentation	Bin Width	25

Original, refers to no prior image normalisation or resegmentation. 19 different image setting configurations were investigated in both single-slice and multi-slice analysis.

4.2.6 Multi-slice: Image Resampling and Interpolation Method

Higher-order radiomic feature calculations require isotropic images, i.e. the pixel size in the *x, y,* and *z* directions should be the same, to be rotationally invariant[189,190]. In CT imaging, images are often isotropic in-plane (*x* and *y* are equal) but will have a larger *z*-axis slice spacing and therefore be anisotropic in 3D. Therefore, in radiomics studies, it is common for images to be isotropically resampled.

We investigated the impact of different image resampling methods on extracted radiomic features: (1) B-spline interpolation and (2) linear interpolation to $1 \times 1 \times 1$ mm^3. Whereas B-spline interpolation is the PyRadiomics default for image resampling, linear interpolation is another method available that tends to be faster.

The execution time to process the 82 carotid arteries in multi-slice analysis using PyRadiomics and the two different image resampling methods was measured from the command line. The tests were conducted on a MacOS X (Catalina, version 10.15.6) computational platform equipped with a 2.8 GHz quad-core Intel Core i7 CPU and 16 GB RAM.

4.2.7 Statistical Analysis

Descriptive statistics are reported as mean (standard deviation [SD]) for continuous variables or median (interquartile range [IQR]) as appropriate.

We measured the degree of robustness using the 2-way mixed-effects model, absolute agreement, single rater and the two-way mixed-effects model, consistency, single rater intraclass correlation coefficient (ICC) according to the McGraw and Wong convention and according to the ICC guidelines of Koo and Li[167], as appropriate.

Let *n* and *k* be the number of subjects and number of raters/measurements, respectively. The ICCs used are defined as follows:

Equation 4.6 Two-way mixed effects, consistency, single rater/measurement

$$ICC(3,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$$

Equation 4.7 Two-way mixed effects, absolute agreement, single rater/measurement

$$ICC(2,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where MS_R, MS_E and MS_C are the mean square for rows, mean square for error and mean square for columns, respectively.

ICC values can range between 0 and 1. In this study, ICC values greater than or equal to 0.9 were considered to have 'excellent' robustness. Radiomic features were classified into three groups, with ICC values < 0.5, between 0.5 to 0.9 and ≥ 0.9, being indicative of poor, moderate and excellent robustness (also referred to as highly robust), respectively[168].

All statistical analysis was performed in IBM SPSS Statistics for Macintosh (Version 25.0. Armonk, NY, USA: IBM Corp.) and Python 3.7.4.

4.2.8 Machine Learning Classification with Highly robust features

4.2.8.1 Non-redundant features

To reduce multicollinearity and feature redundancy, pairwise feature-to-feature correlations were calculated using the Spearman Rank Correlation. For pairs of features with absolute values of Spearman's $|r_s| \geq 0.95$, the features with the highest area under the receiver operating curve (AUC) in univariate logistic regression was retained, and the latter was discarded. The threshold of 0.95 was chosen as this was used previously by Owens et al. in their radiomics analysis of non-small cell lung cancer from CT scans[191]. A threshold of $|r_s| \geq 0.90$ was also investigated as this was used by Shafiq-ul-Hassan in their lung cancer CT radiomic analysis[192].

4.2.8.2 Machine Learning Predictors

The performance of the highly robust radiomic features was evaluated in their ability to classify culprit versus non-culprit carotid arteries. The carotid calcium score was used to provide a baseline for predictive performance allowing comparisons with the other models (the method for calcium quantification was previously described in Chapter 2, section 2.2.3.1). Finally, an integrated model that incorporated both the highly robust radiomic features and carotid calcium was also evaluated to assess the value of combining radiomic and clinical information for identification of the culprit carotid artery.

4.2.8.3 Data Pre-processing and Machine Learning Classifiers

Each predictor was standardised to have a mean of zero and a variance of one prior to machine learning classification. Using a random state of 42[1] for reproducibility, 6 machine learning classifiers were evaluated using their default configurations in scikit-learn version 0.23.1, which are documented in https://scikit-learn.org/stable/tutorial/index.html. The following ML classifiers were chosen as they are commonly used in feature-based machine learning applications: decision tree[193], random forest[194], LASSO regression[195], Elastic Net regression[196], a neural network[197] and XGBoost[198].

The Gini Impurity criterion was used for the decision tree and random forest classifiers. For LASSO regression, the norm used in penalisation was set to 'L1' for the sklearn default logistic regression classifier, whilst it was set to 'elasticnet' with an L1 ratio of 0.5 for Elastic Net regression. The norm used in penalisation refers to the method of regularisation employed to prevent overfitting. L1 regularisation involves shrinking the weights of the predictors in such a way that small weights tend to go to zero. Whilst LASSO regression employs L1 regularisation only, the Elastic Net classifier employs a mixture of L1 and L2 regularisation. L2 regularisation typically leads to weights that are homogenously small. The SAGA solver was used for both these logistic regression-based classifiers. The neural network architecture involved 1 hidden layer with 100 hidden units. The ReLU activation function was used, along with an initial learning rate of 0.001, the Adam optimizer and a maximum number of 200 iterations. The XGBoost classifier was implemented using the default settings provided in the XGBoost Python package version 1.2.0, detailed in https://xgboost.readthedocs.io/en/latest/index.html.

The dataset was shuffled and the average performance (accuracy and AUC) of the classifiers calculated following five-fold stratified cross-validation. Stratified cross-validation ensures that when the data is split into its respective folds, there is an equal number of observations for each category i.e. the same proportion of culprit and non-culprit carotid artery cases in each fold.

[1] For the interested reader, the story of why 42 is used so widely as the random seed for reproducibility can be found here: https://medium.com/@leticia.b/the-story-of-seed-42-874953452b94.

4.2.8.4 Comparing model performances

In five-fold cross-validation, an AUC is provided for the model performance in each fold. The mean cross-validated AUC is the average of the AUC values across the five folds. DeLong's method[199] was used to compare the AUC of the radiomics features-only model and of the integrated models (using radiomic features and carotid calcium as predictors) with the baseline calcium-only model in each fold of the cross-validation scheme for both single- and multi-slice approaches. The distribution of five AUC values across the folds within the cross-validation scheme was compared using the Wilcoxon signed-rank test for the following comparisons: (1) calcium-only versus radiomics features-only model, (2) calcium-only versus integrated model and (3) radiomics-only versus integrated model. A p-value <0.05 was considered statistically significant.

4.3 Results

4.3.1 Robustness analysis carotid dataset

Carotid CTA scans from 41 symptomatic patients were analysed in this robustness analysis, comprising 41 culprit and 41 non-culprit carotid arteries (82 carotid arteries in total). In single-slice analysis, the ROIs were drawn on axial slices with a mean in-plane pixel size of 0.56×0.56 mm and a slice thickness of 0.625 mm with 0.4 mm slice spacing. For multi-slice analysis, CTA scans were resampled to 3 mm slice thickness, resulting in a slice spacing of 1.92 mm before manual segmentation of the carotid artery on 14 consecutive axial slices (see Chapter 2, section 2.2.4.4). Prior to radiomic feature extraction with PyRadiomics in multi-slice analysis, the CTA image and the segmentation mask were further resampled to 1×1×1 mm^3 to extract 3D radiomic features from isotropic voxels (see Chapter 3, section 3.2.2.5).

4.3.2 Segmentation perturbations, intra-observer and inter-observer variabilities

The Dice coefficient was calculated to assess agreement between ROI segmentations in the following ways: (1) comparing the ROIs for 8 carotid arteries drawn by the primary reader at two separate time points to determine intra-observer variability, (2) comparing ROIs for 8 carotid arteries drawn by the primary observer with those drawn by a second independent observer to determine inter-observer variability and (3) comparing the ROIs for 82 carotid arteries drawn by the primary observer with the ROIs generated following morphological operations (dilations and erosions) to determine the variability generated by systematic ROI perturbations.

As demonstrated in Figure 4.49, there was low variability in intra-observer segmentation with a mean Dice coefficient (SD) of 0.95 (0.02), whereby a Dice coefficient value of 1 indicates complete spatial overlap between segmentations. There was greater variability in inter-observer segmentation with a mean DC of 0.85 (0.06). The morphological operations applied to the ROIs captured the range of variability that occurred with human inter-observer variability, with a mean DC of 0.86 (0.06).

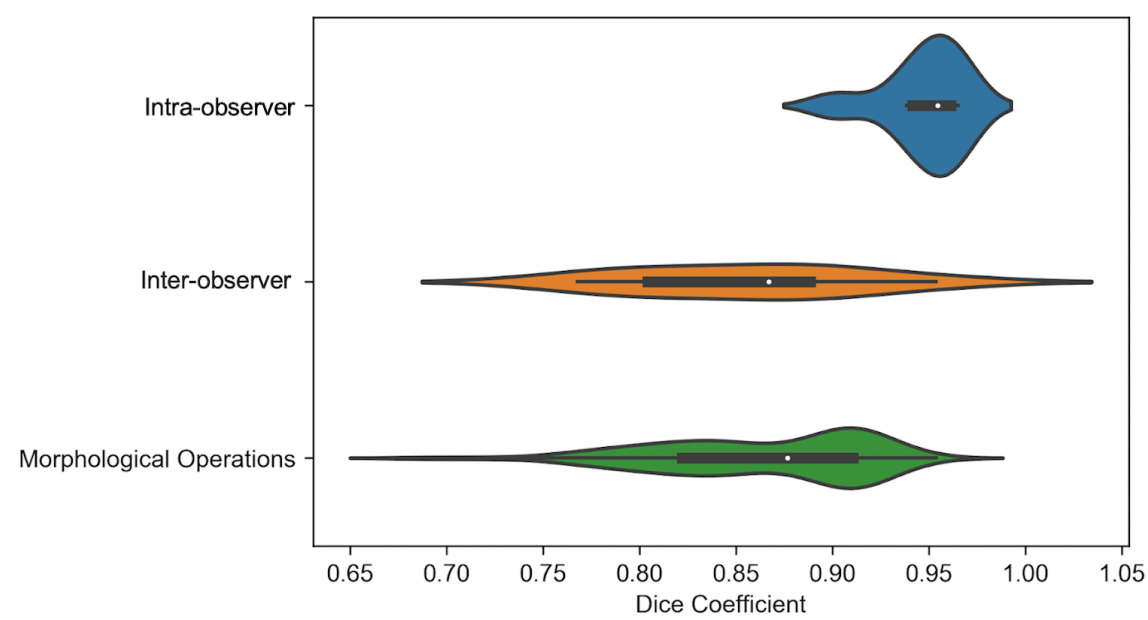


Figure 4.49 Violin plots of image segmentation agreement as determined by the Dice coefficient with the original ROI

Intra-observer variability was determined by manual segmentation of the ROIs by the same observer (EPVL) at two separate time points for 8 carotid arteries. Inter-observer variability was determined by manual ROI segmentation by two independent observers (EPVL and CW), performed on 8 carotid arteries. The ROIs drawn by the primary observer (EPVL) on 82 carotid arteries were compared with those generated following dilation and erosion morphological operations in single-slice analysis to determine the Dice coefficient distribution of systematic ROI perturbations. The violin plots each contain a black box-plot of the Dice coefficients, with the white dot representing the median Dice coefficient. Whereas the box-plots correspond to the actual data points of the Dice coefficients, the coloured shapes provide a visualisation of the underlying Dice coefficient distributions via kernel density estimations.

4.3.3 Proportion of robust radiomic features in different image settings

4.3.3.1 Feature robustness in single-slice analysis

There were 19 different image settings investigated that included different combinations of 1) original image, 2) prior normalisation and 3) resegmentation with different image quantisation configurations using bin widths ranging from 10-35 (in increments of 5) or bin counts ranging from 8-256 (in powers of 2).

4.3.3.1.1 Original Image

Over 50% of the 93 radiomic features had excellent robustness to ROI perturbations when using the original image and a fixed bin width for image quantisation, ranging from 10 to 35. Using a fixed bin width of 10 led to a higher proportion of poorly robust radiomic features compared with the other investigated bin widths. The bin width setting that had the most highly robust features for single-slice analysis was bin width 25 or 30. This corresponds with the PyRadiomics default setting (fixed bin width of 25).

A detailed breakdown of the radiomic feature classes and their corresponding robustness categories (excellent, moderate or poor robustness) according to the ICC (absolute agreement) value for single-slice analysis (original image) and a fixed bin width of 25 is provided in Table 4.65. For this setting, 49 out of 93 (52.7%) radiomic features displayed excellent robustness, 33 (35.5%) had moderate robustness, and 11 (11.8%) were poorly robust. Within the excellent robustness category, the first order feature class had the highest proportion of highly robust radiomic features (61.1%).

There was a small, but non-negligible proportion of radiomic features that were not robust (11.8%) that spanned the spectrum of radiomic feature classes (except the NGTDM feature class). In this setting, the poorly robust features included: First Order: 10^{th} Percentile, First Order: Minimum, GLDM: Low Grey Level Emphasis, GLRLM: Short Run Low Grey Level Emphasis and GLSZM: Small Area Low Grey Level Emphasis.

Table 4.65 Radiomic Feature Class and Robustness: single-slice (original, BW 25)

ICC (absolute agreement)	Number of radiomic features per feature class with:		
Feature Class	**Excellent Robustness ICC ≥ 0.9**	**Moderate Robustness 0.5 ≤ ICC < 0.9**	**Poor Robustness ICC < 0.5**
First Order (n=18)	11 (61.1%)	5 (27.8%)	2 (11.1%)
GLCM (n=24)	14 (58.3%)	9 (37.5%)	1 (4.2%)
GLDM (n=14)	6 (42.9%)	5 (35.7%)	3 (21.4%)
GLRLM (n=16)	9 (56.3%)	4 (25.0%)	3 (18.8%)
GLSZM (n=16)	7 (43.8%)	7 (43.8%)	2 (12.5%)
NGTDM (n=5)	2 (40.0%)	3 (60.0%)	0 (0.0%)
Total (n=93)	49 (52.7%)	33 (35.5%)	11 (11.8%)

Break down of radiomic features by robustness category and feature class type in single-slice analysis (original image and bin width of 25). BW, bin width; ICC, intraclass correlation coefficient; n, number of radiomic features. Frequency of radiomic features per feature class as a number and percentage according to robustness category (excellent, moderate or poor).

When using a fixed bin number, compared with using a fixed bin width for image quantisation, this led to a lower proportion of radiomic features with excellent robustness. Amongst all the fixed bin numbers investigated, 33.3% was the highest proportion of highly robust radiomic features achieved (bin number: 256). The lowest proportion of highly robust radiomic features was 19.4% (bin number: 8 and 16). Although a fixed bin number of 256 had the highest proportion of radiomic features with excellent robustness, it also had the highest proportion of poorly robust radiomic features (15.1%) out of any image quantisation method and value.

4.3.3.1.2 Prior normalisation

Prior normalisation of the image reduced the proportion of poorly robust radiomic features but did not impact the proportion of texture features with excellent robustness, when compared to no prior normalisation (using fixed bin numbers only), see Figure 4.50. This was lower than no prior normalisation and using a fixed bin width however (see section 4.3.3.1.1). The highest proportion of highly robust radiomic features following prior normalisation was 34.4% (bin count: 256).

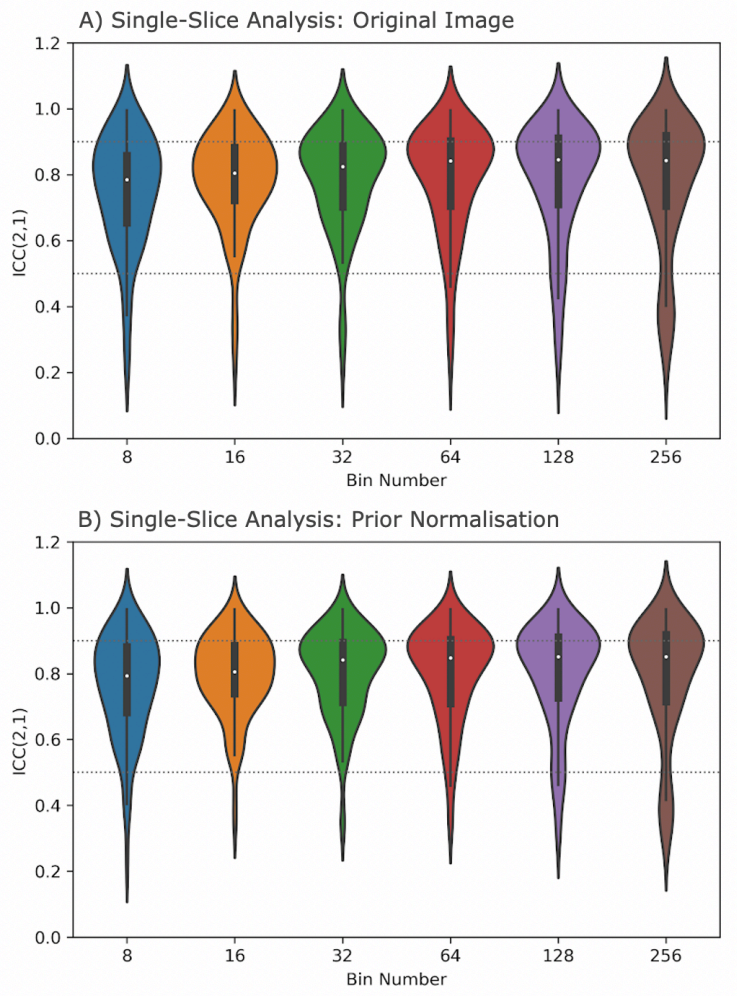


Figure 4.50 Violin plots of ICC (absolute agreement) for robustness: single-slice analysis original image versus prior normalisation

The median ICC(2,1) value is shown as a white dot. Original image refers to no normalisation or resegmentation of the image before radiomic feature extraction. The horizontal line at ICC(2,1)=0.5 denotes the threshold between radiomic features with poor and moderate robustness against ROI perturbations. The horizontal line at ICC(2,1)=0.9 denotes the threshold for a radiomic feature with excellent robustness.

4.3.3.1.3 Resegmentation

Resegmentation reduced the proportion of radiomic features with excellent robustness, but also reduced the proportion of poorly robust texture feature compared with no resegmentation (from 11.8% to 3.2%, bin width: 25). In resegmentation, the majority of radiomic features (88.2%) had moderate robustness (versus 35.5%, original image, bin width: 25). Table 4.66 provides a breakdown of the robustness categories according to radiomic feature class for single-slice analysis with resegmentation.

The radiomic features with excellent robustness were (1) First Order: Energy, (2) First Order: Total Energy, (3) GLDM: Large Dependence High Grey Level Emphasis, (4) GLRLM: Long Run High Grey Level Emphasis, (5) GLSZM: Large Area Emphasis, (6) GLSZM: Large Area High Grey Level Emphasis, (7) GLSZM: Large Area Low Grey Level Emphasis and (8) GLSZM: Zone Variance.

The poorly robust radiomic features were (1) First Order: Minimum, (2) First Order: Range and (3) GLSZM: Grey Level NonUniformity Normalised.

Table 4.66 Radiomic Feature Class and Robustness: single-slice (resegmentation, BW 25)

ICC (absolute agreement)	**Number of radiomic features per feature class with:**		
Feature Class	**Excellent Robustness ICC ≥ 0.9**	**Moderate Robustness 0.5 ≤ ICC < 0.9**	**Poor Robustness ICC < 0.5**
First Order (n=18)	2 (11.1%)	14 (77.8%)	2 (11.1%)
GLCM (n=24)	0 (0.0%)	24 (100.0%)	0 (0.0%)
GLDM (n=14)	1 (7.1%)	13 (92.9%)	0 (0.0%)
GLRLM (n=16)	1 (6.3%)	15 (93.8%)	0 (0.0%)
GLSZM (n=16)	4 (25.0%)	11 (68.8%)	1 (6.3%)
NGTDM (n=5)	0 (0.0%)	5 (100.0%)	0 (0.0%)
Total (n=93)	8 (8.6%)	82 (88.2%)	3 (3.2%)

Resegmentation: Single-slice analysis with fixed bin width of 25

4.3.3.2 Summary of single-slice radiomic features robustness

4.3.3.2.1 Radiomic features with excellent robustness

Figure 4.51 shows the frequency of the different radiomic features that had excellent robustness in the 19 different settings that were investigated. Only 2 texture features (a) GLDM: Large Dependence High Gray Level Emphasis and (b) GLRLM: Long Run High Gray Level Emphasis, demonstrated excellent robustness across all 19 settings (100%). 61 out of 93 (65.6%) extracted texture features demonstrated excellent robustness in at least 1 setting.

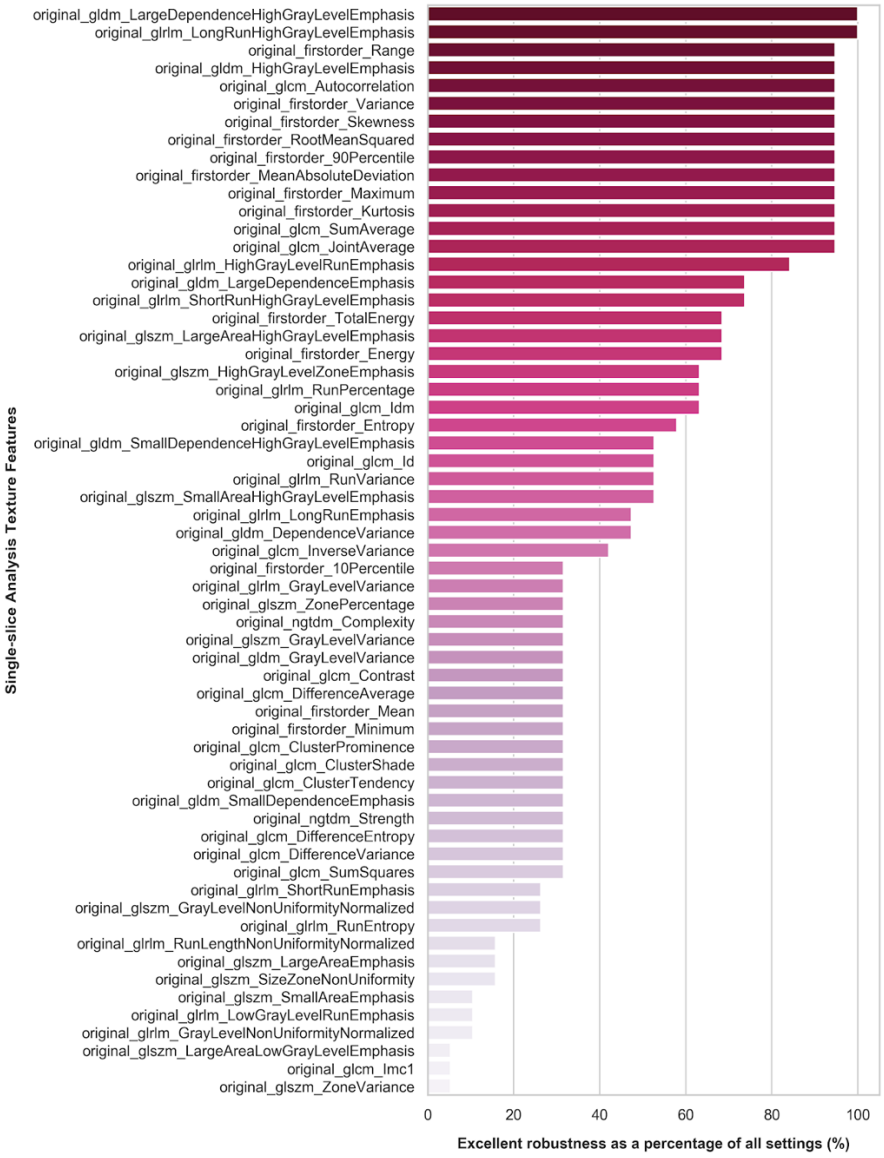


Figure 4.51 Bar chart of single-slice analysis radiomic features with excellent robustness against ROI perturbations across different settings

Excellent robustness was defined as ICC(2,1) ≥ 0.9 comparing the original ROI with ROI dilation (1 and 2 iterations) and ROI erosion (1 and 2 iterations).

4.3.3.2.2 Radiomic features with poor robustness

There were 18 radiomic features that were identified as poorly robust in at least 1 image setting in single-slice analysis. Figure 4.52 provides bar charts in order of ranking for frequency of appearance as a poorly robust radiomic feature amongst the different image settings.

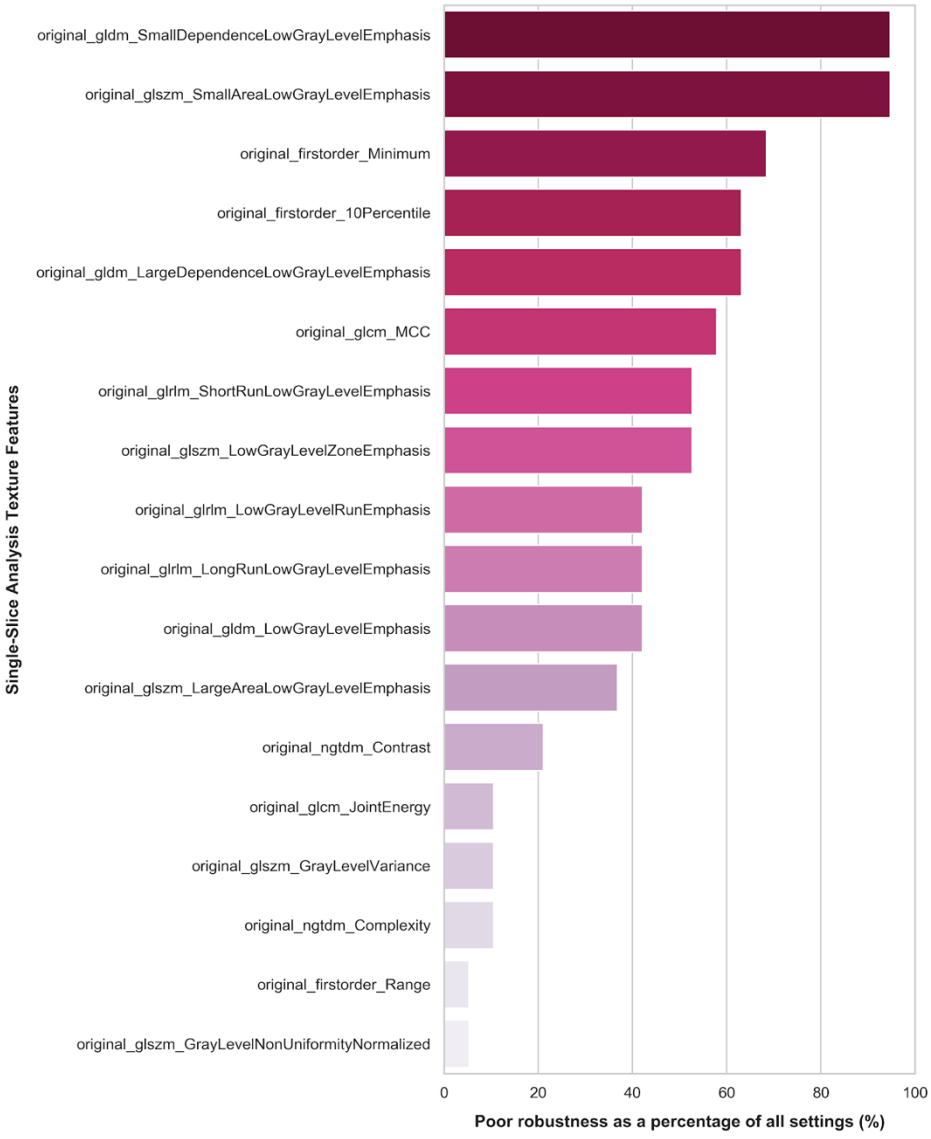


Figure 4.52 Bar chart of single-slice analysis radiomic features that demonstrated poor robustness against ROI perturbations across different image settings

The most frequent radiomic feature with poor robustness amongst the investigated image settings were GLDM: Small Dependence Low Grey Level Emphasis and GLSZM: Small Area Low Grey Level Emphasis which were poorly robust in 18 of the 19 image settings.

4.3.3.3 Feature robustness in multi-slice analysis

4.3.3.3.1 Original Image

In multi-slice analysis, using fixed bin widths, as opposed to fixed bin numbers, led to a higher proportion of highly robust radiomic features, consistent with the single-slice findings. Over 55% of features had excellent robustness when using the original image, with similar proportions across all bin width settings (10-35). Using a fixed bin number rather than a fixed bin width led to a decrease in features with excellent robustness. The highest proportion of highly robust features using a fixed bin number was 48.4% (BN: 256) and the lowest proportion was 28.0% (BN: 8). Table 4.67 provides a breakdown of the feature classes by robustness category in multi-slice analysis (original image) when using the PyRadiomics default of a fixed bin width of 25 and the B-spline interpolation method. In this setting, the number of highly robust features was 52 out of 93 (55.9%) with the greatest proportion coming from the GLCM feature class. The poorly robust features were similar to those identified in single-slice analysis, including First Order: 10th Percentile and GLDM Low Gray Level Emphasis.

Table 4.67 Radiomic Feature Class and Robustness: multi-slice (original, BW 25)

ICC (absolute agreement)	**Number of radiomic features per feature class with:**		
Feature Class	**Excellent Robustness ICC ≥ 0.9**	**Moderate Robustness 0.5 ≤ ICC < 0.9**	**Poor Robustness ICC < 0.5**
First Order (n=18)	10 (55.6%)	6 (33.3%)	2 (11.1%)
GLCM (n=24)	17 (70.8%)	7 (29.2%)	0 (0.0%)
GLDM (n=14)	5 (35.7%)	7 (50.0%)	2 (14.3%)
GLRLM (n=16)	11 (68.8%)	1 (6.3%)	4 (25.0%)
GLSZM (n=16)	6 (37.5%)	8 (50.0%)	2 (12.5%)
NGTDM (n=5)	3 (60.0%)	2 (40.0%)	0 (0.0%)
Total (n=93)	52 (55.9%)	31 (33.3%)	10 (10.8%)

Original image, fixed bin width of 25 and B-spline interpolation used.

4.3.3.3.2 Prior normalisation

Similar to the single-slice analysis findings, prior normalisation in multi-slice analysis reduced the proportion of poorly robust features but had little impact on the proportion of highly robust texture features (when using fixed bin number). The highest proportion of highly robust features after prior normalisation was 49.5% (BN: 256) and the lowest was 29.0% (BN 8).

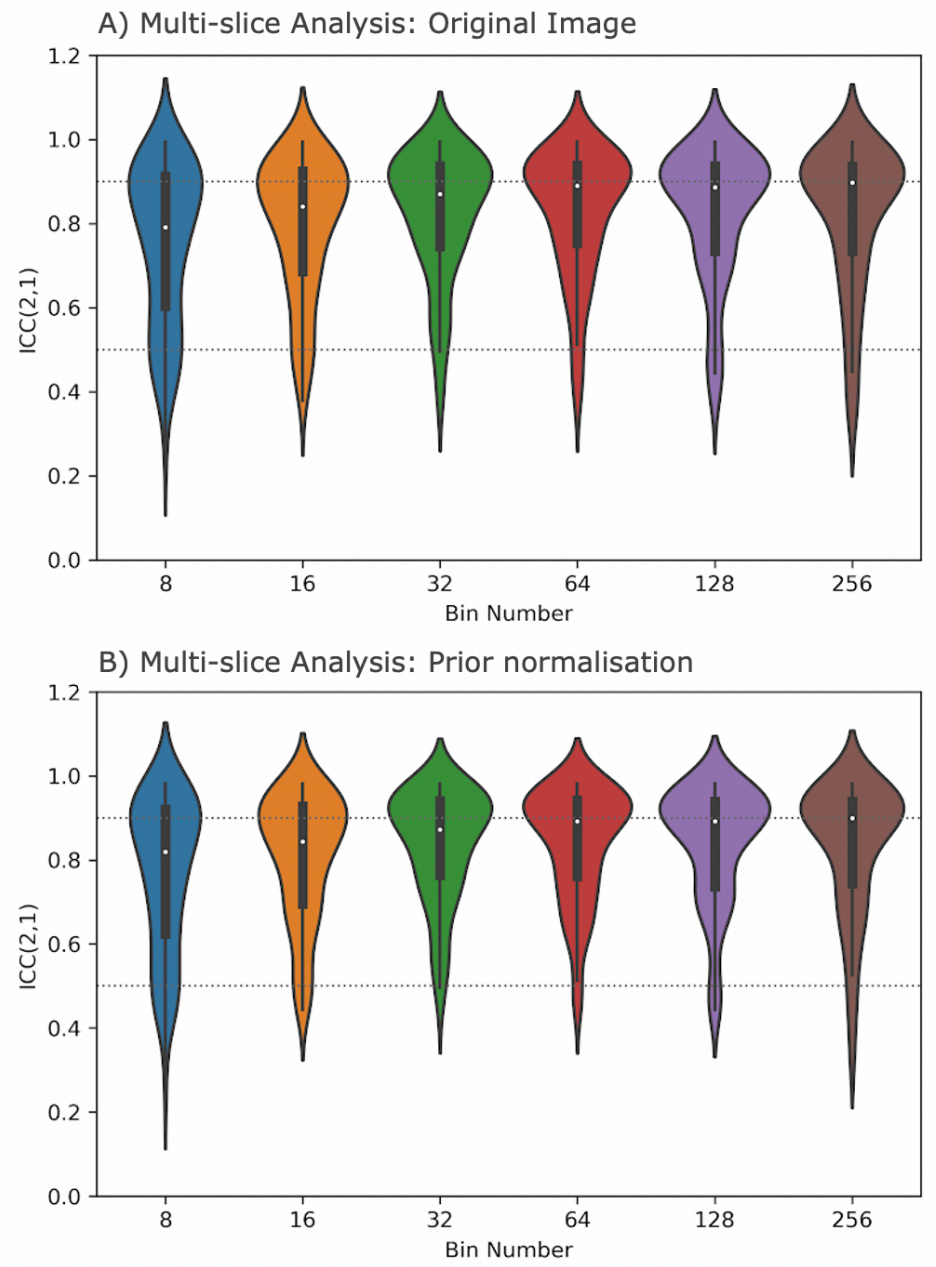


Figure 4.53 Violin plots of ICC (absolute agreement) for robustness against ROI perturbations in multi-slice analysis: original image versus prior normalisation

The median ICC(2,1) value is shown as a white dot. Original image refers to no normalisation or resegmentation. The horizontals line at ICC(2,1)=0.5 denotes the threshold between radiomic features with poor robustness and moderate robustness against ROI perturbations. The horizontal lines at ICC(2,1)=0.9 denotes the threshold for excellent robustness.

4.3.3.3.3 Resegmentation

When resegmentation was performed with a fixed bin width of 25, the proportion of highly robust features decreased (55.9% to 15.1%) when compared with using the original image and the same bin width. However so did the proportion of poorly robust features (10.8% to 2.2%). The majority of texture features shifted to moderate robustness (82.8%).

Table 4.68 Radiomic Feature Class and Robustness: multi-slice (resegmentation, BW 25)

ICC (absolute agreement)	**Number of radiomic features per feature class with:**		
Feature Class	**Excellent Robustness ICC ≥ 0.9**	**Moderate Robustness 0.5 ≤ ICC < 0.9**	**Poor Robustness ICC < 0.5**
First Order (n=18)	3 (16.7%)	13 (72.2%)	2 (11.1%)
GLCM (n=24)	2 (8.3%)	22 (91.7%)	0 (0.0%)
GLDM (n=14)	3 (21.4%)	11 (78.6%)	0 (0.0%)
GLRLM (n=16)	3 (18.8%)	13 (81.3%)	0 (0.0%)
GLSZM (n=16)	3 (18.8%)	13 (81.3%)	0 (0.0%)
NGTDM (n=5)	0 (0.0%)	5 (100.0%)	0 (0.0%)
Total (n=93)	14 (15.1%)	77 (82.8%)	2 (2.2%)

Multi-slice analysis: Resegmentation, fixed bin width = 25, B-spline interpolation. Definition of ICC used = 2-way mixed-effects model, absolute agreement, single rater intraclass correlation coefficient; n, number of radiomics features

In the resegmentation setting, using a bin width of 10 did not change the proportion of radiomic features in each robustness category. There were still 14 features with excellent robustness, 77 with moderate robustness and 2 with poor robustness features. When using a bin width of 15, there were 15 features with excellent robustness, 76 with moderate robustness and 2 with poor robustness features. Using a bin width of 20, there were 14 features with excellent robustness, 77 with moderate and 2 with poor robustness. In each bin width setting (10-25), the two poorly robust features were First Order: Minimum and First Order: Range.

4.3.3.4 Summary of multi-slice radiomic feature robustness

4.3.3.4.1 Radiomic features with excellent robustness

Figure 4.54 shows the frequency of the different radiomic features that showed excellent robustness in the 19 different settings that were investigated in multi-slice analysis. Only 4 features demonstrated excellent robustness across all settings: (1) GLDM Grey Level Variance, (2) First Order Mean Absolute Deviation, (3) GLRLM Grey Level Variance and (4) GLDM: Large Dependence High Grey Level Emphasis. 67 (72%) radiomic features demonstrated excellent robustness in at least 1 setting.

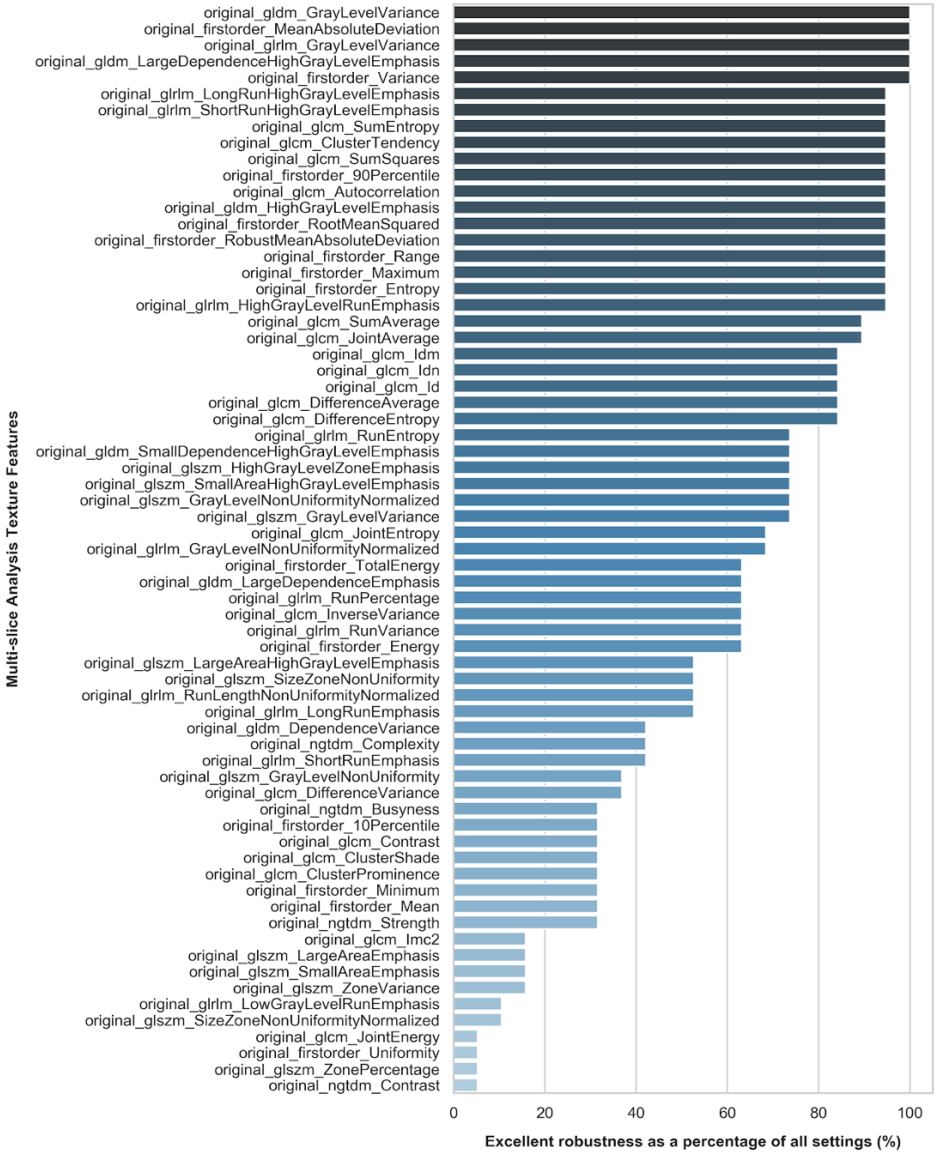


Figure 4.54 Bar chart of multi-slice analysis radiomic features with excellent robustness against ROI perturbations across different settings

Excellent robustness was defined as ICC(2,1) ≥ 0.9 comparing the original ROI with ROI dilation (1 and 2 iterations) and ROI erosion (1 iteration only).

4.3.3.4.2 Radiomic features with poor robustness

In multi-slice analysis, 23 radiomic features were poorly robust in at least one of the investigated image settings. There were no texture features that were poorly robust in all 19 settings. First Order: Minimum was the most common poorly robust radiomic feature. Figure 4.55 presents bar-charts in order of the frequency that radiomic features appeared as poorly robust.

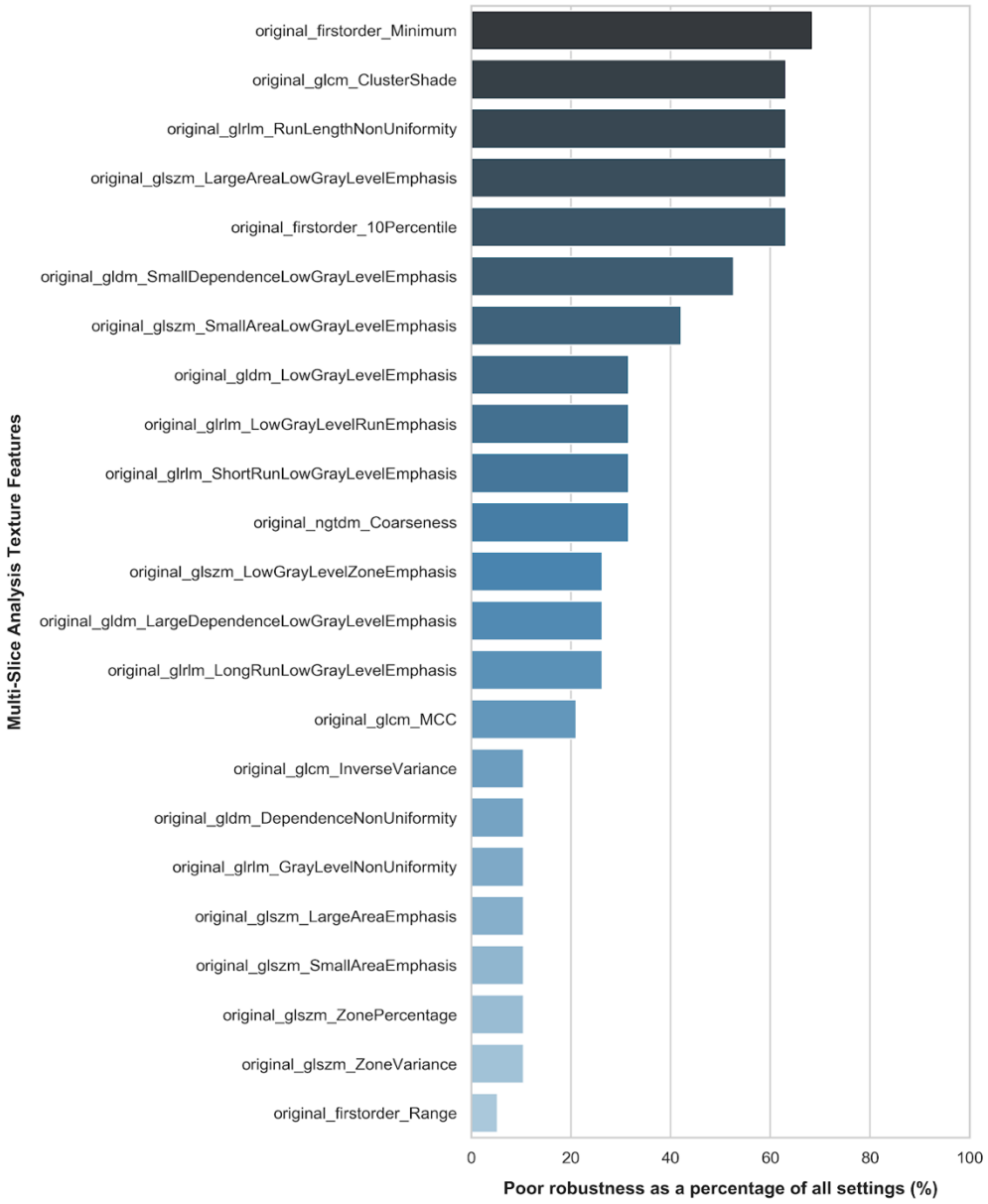


Figure 4.55 Bar chart of multi-slice analysis radiomic features that demonstrated poor robustness against ROI perturbations across different image settings

4.3.3.5 Similarities and differences in single-slice and multi-slice approaches

In single-slice analysis, 61 features had excellent robustness in at least one image setting out of the 19 settings investigated, whilst 67 features had excellent robustness in multi-slice analysis. There was considerable overlap in the radiomic features with excellent robustness between single-slice and multi-slice approaches (n=56), these included First Order: Variance and GLCM: Autocorrelation. However, there were also radiomic features that had excellent robustness in the single-slice approach only (n=5) such as First Order: Kurtosis and GLDM: Small Dependence Emphasis or the multi-slice approach only (n=11) such as First Order: Uniformity and GLCM: Joint Energy, see Figure 4.56A.

In single-slice analysis, 18 features were identified as poorly robust in at least 1 out of the 19 settings investigated, whilst 23 features were identified as poorly robust in multi-slice analysis. There were no radiomic features that were poorly robust in all 19 settings.

Figure 4.56 illustrates the radiomic features that were identified as poorly robust in both single-slice and multi-slice analysis, as well as those unique to only single-slice or multi-slice analysis, respectively. Figure 4.56B illustrates the radiomic features that were identified as poorly robust in both single-slice and multi-slice analysis (n=13) such as First Order: 10th Percentile and GLDM: Low Grey Level Emphasis, as well as those unique to single-slice (n=5) analysis such as GLSZM: Grey Level Variance and NGTDM: Contrast or multi-slice analysis (n=10) such as GLCM: Inverse Variance and NGTDM: Coarseness.

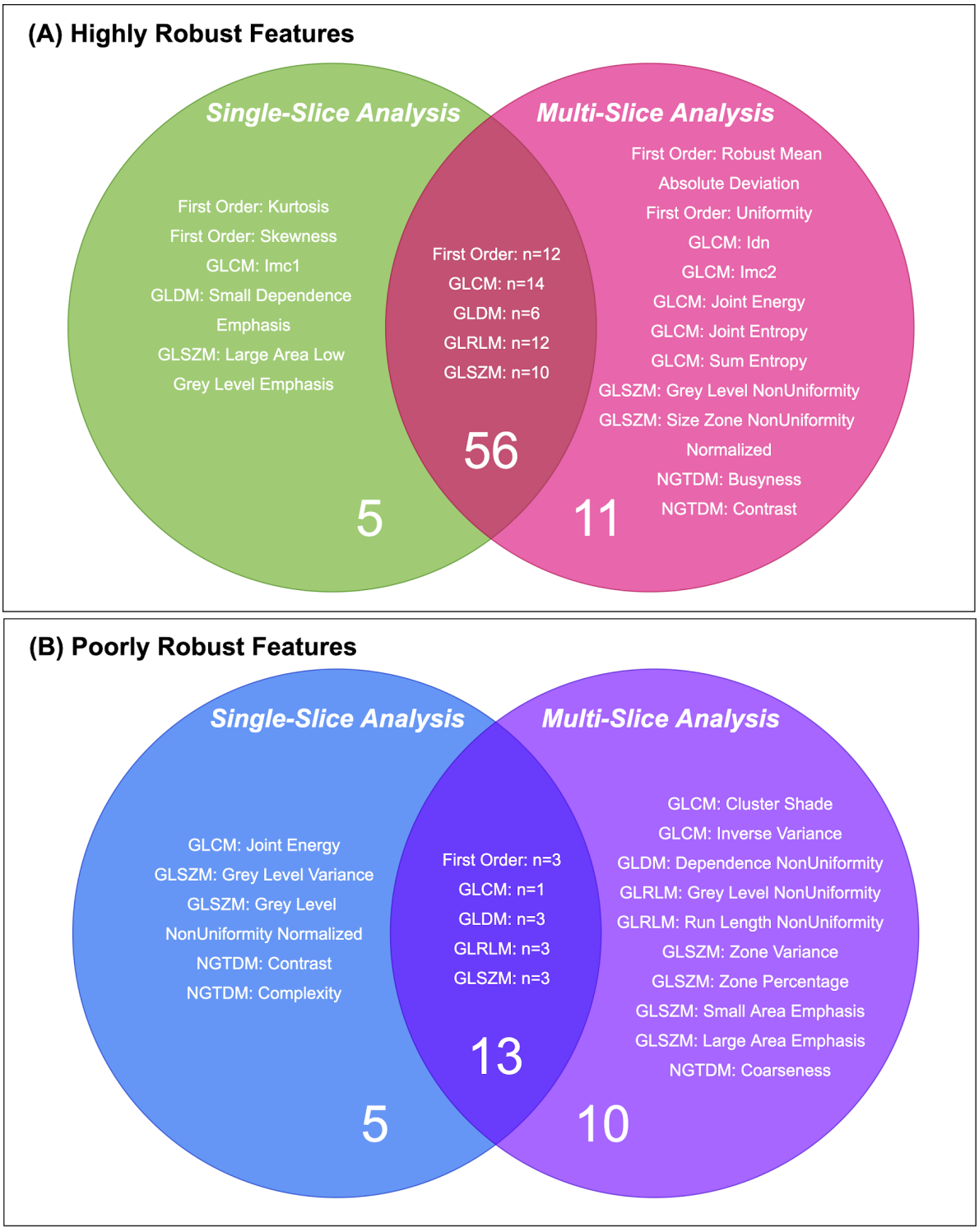


Figure 4.56 Venn Diagram to show radiomic features that were (A) highly robust in single-slice or multi-slice analysis and, (B) poorly robust in single-slice or multi-slice analysis

Amongst the texture features that demonstrated excellent robustness in at least 1 of the 19 image settings, there were 5 features that were robust in single-slice analysis only, and 11 features that were robust in multi-slice analysis only There were poorly robust texture features that occurred in both single-slice and multi-slice analysis, but also in single-slice only and multi-slice only, respectively. The majority of texture features that were poorly robust in both analyses related to low grey levels.

4.3.3.6 Multi-slice: Image Resampling and Interpolation Method

We investigated the effect of using B-spline versus linear interpolation on the extracted texture features and found that the vast majority of texture features show high agreement in both absolute agreement and consistency between the two methods of interpolation. There were no poorly robust features with regards to the method of interpolation. In terms of execution time, for the 82 carotid arteries analysed in multi-slice analysis, when using B-spline interpolation with a fixed bin width of 25, the total time taken was 48.15 seconds. When using linear interpolation with a fixed bin width of 25, the total time taken was 29.49 seconds. The breakdown by feature class is shown in Table 4.69.

Table 4.69 Level of agreement between B-spline and linear interpolation by feature class

	ICC(2,1) absolute agreement			**ICC(3,1) consistency**		
Feature Class	**Excellent Robustness ICC ≥ 0.9**	**Moderate Robustness 0.5 – 0.9**	**Poor Robustness ICC < 0.5**	**Excellent Robustness ICC ≥ 0.9**	**Moderate Robustness 0.5 – 0.9**	**Poor Robustness ICC < 0.5**
First Order (n=18)	18 (100.0%)	0 (0.0%)	0 (0.0%)	18 (100.0%)	0 (0.0%)	0 (0.0%)
GLCM (n=24)	23 (95.8%)	1 (4.2%)	0 (0.0%)	24 (100.0%)	0 (0.0%)	0 (0.0%)
GLDM (n=14)	11 (78.6%)	3 (21.4%)	0 (0.0%)	12 (85.7%)	2 (14.3%)	0 (0.0%)
GLRLM (n=16)	12 (75.0%)	4 (25.0%)	0 (0.0%)	13 (81.3%)	3 (18.8%)	0 (0.0%)
GLSZM (n=16)	13 (81.3%)	3 (18.8%)	0 (0.0%)	13 (81.3%)	3 (18.8%)	0 (0.0%)
NGTDM (n=5)	5 (100.0%)	0 (0.0%)	0 (0.0%)	5 (100.0%)	0 (0.0%)	0 (0.0%)
Total (n=93)	82 (88.2%)	11 (11.8%)	0 (0.0%)	85 (91.4%)	8 (8.6%)	0 (0.0%)

Image type: multi-slice analysis, original (no normalisation/resegmentation), fixed bin width 25

4.3.4 Machine Learning Classification with Highly Robust Features
4.3.4.1 Feature redundancy

We used the Spearman rank correlation to assess multicollinearity between the different pairs of radiomic features, with a threshold of $|r_s| > 0.95$ as the threshold for redundant features.

Table 4.70 Non-redundant radiomic features with excellent robustness in different image settings with threshold of 0.95

Single-slice (Bin width=25)	
Original (n=14)	**Resegmentation (n=7)**
1. First Order: Entropy 2. First Order: Range 3. First Order: Kurtosis 4. First Order: Skewness 5. First Order: Total Energy 6. GLCM: Joint Average 7. GLCM: Cluster Shade 8. GLCM: Inverse Variance 9. GLDM: Large Dependence High Grey Level Emphasis 10. GLRLM: Run Entropy 11. GLRLM: Long Run High Grey Level Emphasis 12. GLRLM: Run Percentage 13. GLSZM: Large Area High Grey Level Emphasis 14. GLSZM: Size Zone NonUniformity	1. First Order: Energy 2. First Order: Total Energy 3. GLDM: Large Dependence High Grey Level Emphasis 4. GLRLM: Long Run High Grey Level Emphasis 5. GLSZM: Large Area Emphasis 6. GLSZM: Large Area High Grey Level Emphasis 7. GLSZM: Large Area Low Grey Level Emphasis
Multi-slice (B-spline interpolation, Bin width=25)	
Original (n=14)	**Resegmentation (n=10)**
1. First Order: 90th Percentile 2. First Order: Root Mean Squared 3. First Order: Maximum 4. First Order: Robust Mean Absolute Deviation 5. GLCM: Cluster Shade 6. GLCM: Cluster Tendency 7. GLCM: Inverse Variance 8. GLCM: Joint Average 9. GLDM: LargeDependenceEmphasis 10. GLDM: Large Dependence High Grey Level Emphasis 11. GLSZM: Small Area Emphasis 12. GLSZM: Size Zone NonUniformity 13. NGTDM: Busyness 14. NGTDM: Complexity	1. First Order: Mean Absolute Deviation 2. GLCM: Difference Variance 3. GLCM: Joint Energy 4. GLDM: Dependence Variance 5. GLDM: Large Dependence High Grey Level Emphasis 6. GLRLM: Long Run High Grey Level Emphasis 7. GLRLM: Grey Level NonUniformity Normalised 8. GLSZM: Grey Level NonUniformity 9. GLSZM: Large Area High Grey Level Emphasis 10. GLSZM: Size Zone NonUniformity

A threshold of $|r_s| > 0.90$ was also investigated, and the non-redundant features with excellent robustness identified for the different image settings is shown in Table 4.71.

Table 4.71 Non-redundant features radiomic features with excellent robustness in different image settings with threshold of 0.9

Single-slice (Bin width=25)	
Original (n=10)	**Resegmentation (n=6)**
1. GLCM: Cluster Shade 2. GLRLM: Run Percentage 3. First Order: Skewness 4. GLDM: Large Dependence High Grey Level Emphasis 5. GLRLM: Run Entropy 6. GLCM: Inverse Variance 7. GLRLM: Long Run High Grey Level Emphasis 8. First Order: Total Energy 9. GLSZM: Large Area High Grey Level Emphasis 10. GLCM: Contrast	1. First Order: Energy 2. First Order: Total Energy 3. GLDM: Large Dependence High Grey Level Emphasis 4. GLRLM: Long Run High Grey Level Emphasis 5. GLSZM: Large Area Emphasis 6. GLSZM: Large Area Low Grey Level Emphasis
Multi-slice (B-spline interpolation, Bin width=25)	
Original (n=9)	**Resegmentation (n=8)**
1. GLCM: Inverse Variance 2. GLCM: Cluster Shade 3. GLCM: Joint Average 4. NGTDM: Busyness 5. GLCM: Idn 6. GLDM: Large Dependence High Grey Level Emphasis 7. GLDM: Large Dependence Emphasis 8. First Order: Variance 9. First Order: Robust Mean Absolute Deviation	1. GLCM: Difference Variance 2. GLDM: Dependence Variance 3. GLCM: Joint Energy 4. First Order: Mean Absolute Deviation 5. GLSZM: Size Zone NonUniformity 6. GLSZM: Grey Level NonUniformity 7. GLRLM: Long Run High Grey Level Emphasis 8. GLDM: Large Dependence High Grey Level Emphasis

4.3.4.2 Culprit versus non-culprit carotid arteries: Binary classification
4.3.4.2.1 Machine learning performance

Since a fixed bin width of 25 was found to produce the highest proportion of radiomic features with excellent robustness, we used the radiomic features extracted using this image quantisation level in the following image settings for machine learning classification: (1) single-slice approach: original image, (2) single-slice approach: with resegmentation, (3) multi-slice approach: original image and (4) multi-slice approach: with resegmentation. Different sets of non-redundant radiomic features with excellent robustness were identified depending on the image setting used (1-4), these radiomic features were detailed in Table 4.70.

Several machine learning classifiers were investigated in a five-fold cross-validation scheme using (1) carotid calcium score as the only predictor, (2) radiomic features (non-redundant with excellent robustness) as the only predictors and (3) radiomic features with carotid calcium. Score (termed the 'integrated model') as predictors to differentiate culprit from non-culprit carotid arteries.

When using the radiomic features sets identified using a threshold of $|r_s| > 0.95$, the image setting that led to the highest predictive performance was the multi-slice approach with resegmentation (image setting 4). Within this setting, the best performing machine learning classifier amongst those investigated was the Elastic Net logistic regression-based classifier. Using this classifier, carotid calcium score alone was a poor predictor of culprit versus non-culprit carotid artery status, performing worse than chance. The mean AUC for carotid calcium score alone was 0.44 (0.11) and the mean accuracy was 46% (7.5%). Using radiomic features alone as predictors performed better than carotid calcium alone, with a mean AUC of 0.67 (0.08) and a p-value of 0.043. The combination of radiomic features with carotid calcium as predictors led to the highest predictive performance, with a mean AUC of 0.73 (0.09, a mean accuracy of 69% (6.0%) and a p-value of 0.043 when compared with carotid calcium alone and a p-value of 0.042 when compared with radiomic features alone. The performance (mean AUC with SD) of the other machine learning classifiers using radiomic features and carotid calcium as predictors were: decision tree 0.58 (0.19), random forest 0.67 (0.08), LASSO 0.72 (0.09), neural network 0.60 (0.09) and XGBoost 0.56 (0.09). Figure 4.57 demonstrates the receiver operating characteristic curves for the Elastic Net classifier and the different predictors.

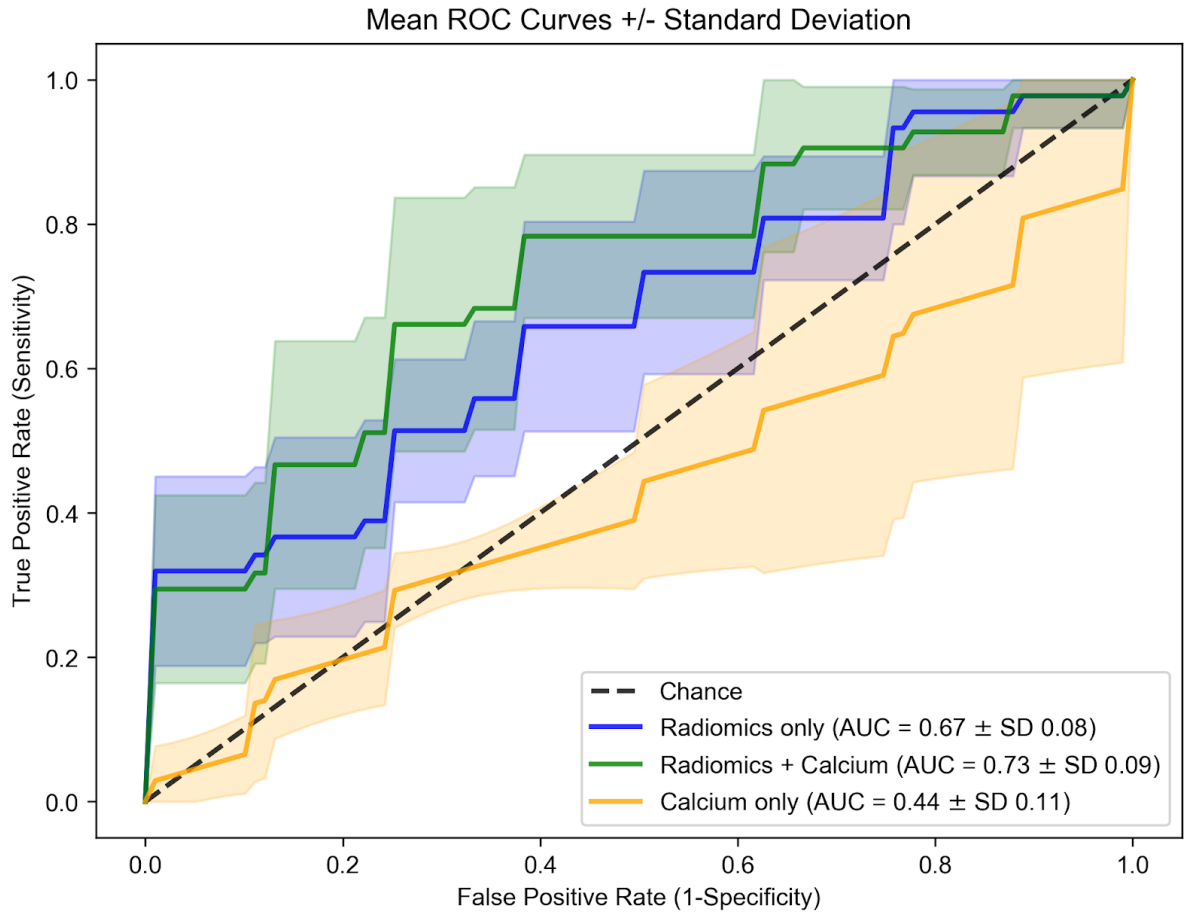


Figure 4.57 Mean receiver operating characteristic (ROC) curves of five-fold stratified cross-validation in multi-slice analysis with resegmentation for Elastic Net logistic regression classifier

Elastic Net classifier (weight L_1 and L_2 penalties=0.5) using (1) radiomic features only as predictors, (2) radiomic features and calcium and (3) calcium only. Image setting: multi-slice analysis with resegmentation and a fixed bin width of 25. Dashed line indicates expected AUC for a random chance classifier. AUC, area under the ROC curve; SD, standard deviation.

Whilst the combination of carotid calcium and radiomic features led to improved predictive performance versus either carotid calcium alone or radiomic features alone as predictors in multi-slice analysis with resegmentation, this was not the case for the other image settings. The performance of Elastic Net in the other image settings was as follows: (1) single-slice analysis (original image), radiomics-only model: AUC 0.47 (0.13) and accuracy 51.1% (SD 10%); integrated model: AUC 0.44 (0.12) and accuracy 42.5% (7%), (2) single-slice analysis (with resegmentation): radiomics-only model: AUC 0.55 (0.13) and accuracy 51.4% (9%), integrated model: AUC 0.54 (0.14) and accuracy 48.9% (SD 0.07) and (3) multi-slice analysis (original image), radiomics-only and integrated model: AUC 0.67 (0.10) and accuracy 62.6% (12%).

When using the radiomic features sets identified using a threshold of $|r_s| > 0.90$, the image setting that led to the highest predictive performance was the multi-slice approach with resegmentation (image setting 4). Within this setting, the best performing machine learning classifier amongst those investigated was the LASSO classifier according to AUC and the Elastic Net classifier according to accuracy. Using the integrated model, the mean AUC for the LASSO classifier was 0.67 (0.15). and the mean accuracy was 58.5% (7%). For the Elastic Net classifier, the mean AUC was 0.66 (0.15) and the mean accuracy was 59.7% (6%). Therefore, using a threshold of $|r_s| > 0.95$ rather than 0.90 led to a higher predictive performance.

4.3.4.2.2 Machine learning interpretability

Using a threshold of $|r_s| > 0.95$, the radiomic feature set (n=10) for multi-slice analysis with resegmentation is shown in Figure 4.58 along with the coefficients for each feature as determined by the Elastic Net classifier per fold within the cross-validation scheme.

The feature coefficients indicate the importance of the features for the model's predictions. Larger positive coefficient values suggest higher importance for predicting the culprit carotid artery class, whilst larger negative coefficient values suggest higher importance for predicting the non-culprit carotid artery class.

Overall, there were 3 radiomic features that were highly consistent in being relevant predictors for carotid artery status across every cross-validation fold: (1) GLDM: Dependence Variance, (2) GLSZM: Grey Level NonUniformity and (3) GLRLM: Long Run High Grey Level Emphasis.

	Cross Validation Fold					
	1	2	3	4	5	
	(A) Calcium Only					
AUC	0.37	0.5	0.55	0.5	0.25	**Mean (SD): 0.44 (0.11)**
Radiomic Feature	Feature Coefficients: (B) Radiomic Features Only					**Feature Class**
Dependence Variance	0.746	0.734	0.611	0.766	0.381	GLDM
Size Zone NonUniformity	0.029	0.283	0.798	0.426	0.541	GLSZM
Grey Level NonUniformity	-0.508	-0.930	-0.896	-0.395	-0.588	GLSZM
Long Run High Grey Level Emphasis	-1.077	-0.713	-0.859	-0.766	-0.993	GLRLM
Difference Variance	-0.960	0	-0.601	-0.580	-0.772	GLCM
Grey Level NonUniformity Normalized	-0.162	0	-0.151	-0.080	0	GLRLM
Large Area High Grey Level Emphasis	0	0	0	-0.034	0.017	GLSZM
Mean Absolute Deviation	-0.030	0	0	0	0.005	First Order
Large Dependence High Grey Level Emphasis	0	0.132	0	0	0	GLDM
Joint Energy	0.033	0	0	0	0	GLCM
AUC	0.59	0.56	0.73	0.77	0.69	**Mean (SD): 0.67 (0.08)**
***p*-value**	0.074	0.690	0.247	0.033*	0.088	
Radiomic Feature	Feature Coefficients: (C) Radiomic Features & Calcium					**Feature Class**
Dependence Variance	0.864	0.999	0.812	0.847	0.573	GLDM
Calcium	0.396	0.687	0.648	0.478	0.642	Clinical
Grey Level NonUniformity	-0.508	-0.915	-0.750	-0.256	-0.479	GLSZM
Long Run High Grey Level Emphasis	-1.114	-0.769	-1.024	-0.881	-1.172	GLRLM
Size Zone NonUniformity	-0.033	0	0.328	0.054	0.072	GLSZM
Difference Variance	-1.008	0	-0.604	-0.612	-0.808	GLCM
Grey Level NonUniformity Normalized	0	0.061	-0.065	0	0	GLRLM
Large Area High Grey Level Emphasis	0	0	0	-0.009	0.022	GLSZM
Large Dependence High Grey Level Emphasis	0	0.070	0	0	0	GLDM
Joint Energy	0.055	0	0	0	0	GLCM
Mean Absolute Deviation	0	-0.080	0	0	0	First Order
AUC	0.63	0.63	0.78	0.84	0.75	**Mean (SD): 0.73 (0.09)**
***p*-value**	0.054	0.403	0.065	9.83×10^{-4}***	0.020*	

Figure 4.58 AUC and feature coefficients of predictors used in the Elastic Net logistic regression classifier in multi-slice analysis with resegmentation

*The coefficients of the non-redundant radiomic features with excellent robustness in multi-slice analysis with resegmentation according to the Elastic Net model per fold of the five-fold cross-validation scheme are provided. The name of these radiomic features are provided on the left-hand side, whilst their corresponding feature classes are given on the right-hand side. The AUC for each fold of cross-validation is provided and the mean AUC and standard deviation is stated in bold. The predictive performance of (A) when calcium is used as the only predictor, (B) when only radiomic features are the predictors and (C) when using both radiomic features and calcium in an integrated model, are provided. * p-value<0.05, *** p-value<0.001 when comparing the classification performance per fold of radiomic models B and C with the calcium only model (A) using DeLong's method. The colours highlight non-zero feature coefficients where non-zero coefficients indicate how features played a role in the model's predictions.*

4.4 Discussion

4.4.1 Summary of key findings

4.4.1.1 Summary of key findings: Radiomic feature robustness in different image settings

This chapter focused on characterising radiomic features that were robust to ROI perturbations. These perturbations were used to mimic the ROI placement over- and under-estimations introduced by human subjectivity in clinical practice when using manual delineation methods. The results showed that not all of the 93 extracted radiomic features were robust against these perturbations and the proportion of highly robust radiomic features was dependent on the image setting used. Furthermore, radiomic features identified as highly robust in one setting was not necessarily robust in another setting. There were both similarities and differences in highly robust and poorly robust radiomic features between single-slice analysis and multi-slice analysis.

Multi-slice analysis produced a higher proportion of radiomic features with excellent robustness than single-slice analysis. Using the original image, rather than prior normalisation or resegmentation, along with a fixed bin width (rather than fixed bin number) led to the greatest proportions of highly robust carotid CTA radiomic features. B-spline interpolation and linear interpolation methods for resampling led to similar results in terms of highly robust features.

4.4.1.2 Summary of key findings: Machine learning classification

Although resegmentation reduced the number of radiomic features with excellent robustness and shifted them into the moderate robustness category, this image setting improved the predictive performance of machine learning classifiers for the differentiation of culprit versus non-culprit carotid arteries. The radiomic features set identified as highly robust and non-redundant for the multi-slice approach with resegmentation led to the best predictive model using an Elastic Net classifier with this set of radiomic features alongside carotid calcification as predictors. However, carotid calcification was a poor univariable predictor and in the other settings (single-slice analysis: original image/with resegmentation and multi-slice analysis: original image), the incorporation of carotid calcium with the radiomic features either made no difference or decreased the performance of the machine learning classifiers.

4.4.2 Interpretation of findings

4.4.2.1 Interpretation of findings: Radiomic Feature Robustness in different image settings

There were similarities between the radiomic features with poor robustness in both single-slice and multi-slice analysis to include the radiomic features: First Order: 10th Percentile and GLDM: Low Grey Level Emphasis. These features are related to low grey levels, which in this case reflects low CT attenuation values. This likely reflects the impact of how much perivascular fat surrounding the carotid vasculature is captured in the ROI or VOI segmentation following the morphological perturbations. Indeed, for the human operator manually delineating ROIs, distinguishing the boundary between the carotid artery adventitia and the extra-vessel matrix was a subjective process and likely one of the main drivers for intra- and inter-observer variability. Therefore, identifying the radiomic features that were robust to these perturbations is highly clinically relevant.

Although there was a considerable overlap between the radiomic features with excellent robustness in both single-slice and multi-slice approaches, there were certain radiomic features that had excellent robustness only in either single-slice or multi-slice analysis. This was also the case for radiomic features with poor robustness, for example GLCM: Joint Energy had excellent robustness in multi-slice analysis but poor robustness in single-slice analysis. This may relate to how radiomic features extracted in single-slice analysis would be heavily dependent on ROI placement around the carotid bifurcation only, whereas radiomic features extracted in multi-slice analysis considers more voxels, capturing more information about the carotid artery.

In addition, multi-slice analysis led to a higher proportion of robust radiomic features and had better predictive value than the single-slice approach. These results echo those of oncological radiomic studies which found that using a multi-slice approach compared to a single-slice approach (i.e. whole tumour analysis versus largest cross-sectional area) was more representative of tumour heterogeneity[164] and those of the Chapter 2 and 3, whereby a multi-slice approach revealed further differences between carotid arteries than a single-slice approach.

In this study, prior normalisation of CTA scans with PyRadiomics did not increase the proportion of radiomic features with excellent robustness. Using no prior normalisation and a fixed bin width consistently led to a higher proportion of highly robust radiomic features. Therefore, whereas prior image normalisation is necessary for image pre-processing of MRI scans (since the grey values are arbitrary), it does not seem necessary for carotid CTA scans (where the grey values are calibrated to Hounsfield units). This is in line with most CT imaging radiomic studies which do not tend to apply prior normalisation[169].

Concerning image quantisation methods and values, the use of bin number versus bin width were not interchangeable. Using a fixed bin width between 10-35 consistently led to higher proportions of highly robust radiomic features compared with using a fixed bin number. However, we found a decrease in the proportion of radiomic features with excellent robustness when using the higher limit of 35 and so did not investigate higher values than this limit. Leijenaar et al. investigated different discretisation methods for FDG-PET radiomics analysis of lung tumours and found similar results, with radiomic feature values being dependent on the image quantisation method used and with a fixed bin width being the preferred method[200].

4.4.2.2 Interpretation of findings: Machine Learning Classification

Grey value range resegmentation improved the predictive performance of machine learning classifiers for the differentiation of culprit versus non-culprit carotid arteries. This may reflect how resegmentation excluded high grey values which are related to carotid calcification and luminal contrast, so that differences between culprit and non-culprit carotid artery radiomic profiles could be more easily identified, as seen in Chapter 2 and 3. However, resegmentation reduced the number of radiomic features with excellent robustness and shifted them into a moderate robustness category. On the other hand, resegmentation also reduced the proportion of poorly robust radiomic features, as the low grey values that reflect perivascular fat were excluded. Nevertheless, the radiomic feature set identified for multi-slice analysis with resegmentation contained sufficient information to perform better than carotid calcification alone in identifying the culprit carotid artery. It is not surprising that the carotid calcium score was a poor predictor on its own since there was no statistically significant difference between the calcium burden of culprit versus non-culprit carotid arteries.

In the image settings without resegmentation (single-slice and multi-slice: original image), using the radiomics features together with carotid calcium as predictors did not improve the predictive performance of the models compared with using the radiomic features alone. This is because the radiomic features such as GLDM: Large Dependence High Grey Level Emphasis when extracted from the original image will already be reflecting the carotid calcium score. However, in multi-slice analysis with resegmentation, the integrated model performed better than the radiomic features-only model. This relates to how resegmentation would have limited the effect of calcification on the radiomic features, such that the addition of carotid calcification into the model would provide new information.

4.4.3 Recommendations

In conclusion, we recommend the following be considered for future radiomic studies:

- Since a fixed bin width of 25 or 30 for image quantisation led to the highest proportion of robust radiomic features and the Pyradiomics default is a bin width of 25, we recommend this setting.
- Multi-slice analysis leads to a higher proportion of robust radiomic features than single-slice analysis and leads to a better predictive performance.
- Prior image normalisation is not necessary in carotid CTA radiomic studies and resegmentation may be considered to limit the effects of calcification and luminal contrast for disease characterisation.
- In multi-slice analysis, B-spline or linear interpolation methods led to similar results. Since the PyRadiomics default is already B-spline interpolation, we recommend this setting.
- Always state image settings used in radiomics studies clearly, to include image acquisition protocols, image pre-processing details and method of interpolation, method and value of image quantisation and texture feature definitions.

4.4.4 Limitations

All the images within the robustness analysis carotid imaging dataset were captured using the same scanner, in a single centre (Addenbrooke's Hospital) with similar patient characteristics. Consequently, we could not investigate the effects of inter-scanner variability on the robustness of radiomic features. A quantitative imaging biomarker should be reproducible over time, however in this study we did not have repeat imaging for the patients and so were unable to assess the test-retest repeatability of the radiomic features. This is an area that should be addressed in future prospective studies. It may be difficult to justify imaging human patients at multiple time points for monitoring of the carotid arteries compared with oncological applications and so using a phantom with the characteristics of carotid vasculature may be a solution to assess test-retest repeatability. The phantom can be scanned twice (or more often) within a time interval using the same scanner, acquisition and reconstruction protocol. In addition, phantom studies conducted using different scanners would allow for the assessment of radiomic feature robustness and vendor dependency as well as other considerations within the radiomics workflow such as image reconstruction parameters. Nevertheless, in this study, we aimed to provide a comprehensive robustness analysis of variations that could occur in the radiomics workflow of a single centre, from ROI delineations to image quantisation and image pre-processing methods.

In this study, we have only investigated unfiltered radiomic features. There are other radiomic features that could be extracted following image filtering such as Gabor filters or wavelet transformations. However, as this was a proof-of-principle study using first order and higher-radiomic features, we wanted to limit the number of features extracted. Future work could look into the extraction of further radiomic features following image filtering and their impact on radiomic feature robustness. As the primary objective of this study focused on robustness analyses rather than developing a definitive radiomics signature, the default Python scikit-learn configurations for the machine learning classifiers were used, as opposed to extensive hyperparameter tuning. This avoided further reduction of the limited dataset that could be used for training the machine learning classifiers. Nonetheless, this suggests that the predictive performance we have already achieved, could be further improved with hyperparameter tuning in future work.

Another consideration is that the multi-slice analysis procedure involved resampling the CTA scans from the original 0.625mm slice thicknesses to 3mm slice thickness using the software OsiriX MD, to ensure that the area-of-interest for the carotid artery analysed corresponded to previous PET-CT studies (the rationale is further explained in Chapter 2, section 2.2.4.2). Subsequently, the scans were further resampled to $1\times1\times1$ mm^3 to extract 3D radiomic features from isotropic voxels with PyRadiomics. Therefore, there are two steps of interpolation in the multi-slice analysis workflow. However, this was consistent for all scans in the study.

Lastly, to conclusively assess whether using highly robust radiomic features in predictive modelling improves the models' generalisability, we require testing on external validation data. However, it is difficult to obtain external validation data that acquired CTA scans using the same scanner type, image acquisition protocols and image reconstruction parameters. The lack of standardisation regarding data acquisition to include imaging protocols, scanner types and slice thicknesses is likely to remain a barrier to large radiomics datasets. Therefore, future work could investigate the use of a posteriori harmonisation methods such as ComBat harmonisation[201]. This method originates from those used to overcome batch effects in microarray expression data and has been applied to multicentre radiomic studies in PET imaging[201].

4.5 Conclusions

In this study, a systematic approach was taken to evaluate the robustness of radiomics features against segmentation perturbations in different image settings. Our results demonstrated that (1) not all radiomic features are robust to segmentation perturbations, (2) the set of robust radiomic features is dependent upon the image setting used (including image pre-processing and image quantisation decisions), (3) a multi-slice approach is better than a single-slice approach in terms of radiomic feature robustness and predictive accuracy, (4) there is no need for image normalisation in carotid CTA radiomic studies and (5) grey value range resegmentation can help improve predictive accuracy. In addition, we identified highly robust and non-redundant radiomic features with superior predictive performance to carotid calcification alone in the classification of culprit versus non-culprit carotid arteries of symptomatic patients.

Chapter 5: Multi-class Feature-based Machine Learning with Biological Correlates

Chapter summary:

The previous chapters have established that multi-slice analysis leads to more robust radiomic features and is more predictive than features derived from a single slice of carotid artery alone. In this chapter, we investigate the ability of the previously identified (Chapter 4) highly robust and non-redundant features to classify carotid arteries into 3 possible classes: culprit, non-culprit or asymptomatic. We also investigate the possible biological basis of carotid artery radiomic features.

5.1 Background

A recognised challenge with the quantitative image data mining approach of radiomics is that variations in certain radiomic features may not reflect variations in underlying biology, but rather be due to factors such as changes in image acquisition or during other stages of the radiomics workflow[62]. Previously in Chapter 4, we identified radiomic features that were robust to variations in image segmentations, bringing us one step closer to identifying image-based features that reflect clinically relevant differences between culprit, non-culprit and asymptomatic carotid arteries. In many cases, the subjects in this study had vascular positron emission tomography (PET) performed at the same time as CTA. PET is a non-invasive molecular imaging technique that uses radio-labelled probes that can target specific metabolic processes *in vivo*. This can be used to probe the activity of disease processes, most notably in oncology but also in vascular disease. For example, PET with ^{18}F-fluorodeoxyglucose (FDG) has demonstrated utility in measuring arterial inflammation[202], important for predicting future vascular events and identifying culprit carotid arteries[202,203]. However, FDG is a non-specific measure of glycolytic activity. More specific PET tracers have been tested, such as ^{68}Ga-DOTATATE, targeting activated macrophages present within atherosclerotic plaques[152] and ^{18}F-sodium fluoride (NaF) which is specific for active microcalcification within atherosclerosis[151]. Recently, Evans et al. found that both FDG and NaF uptake were increased in culprit carotid plaques compared with non-culprit plaques in 26 patients following acute ischaemic stroke[151]. In a subset of patients, histology with %CD68+ immunohistochemistry (a marker of macrophage presence) is available from carotid endarterectomy specimens.

Together, having both CT radiomic and PET data (and in some cases histology) on the same subjects provides an opportunity to assess the correlation of carotid CTA radiomic features with specific pathological processes within the artery and may thus provide insights into the biological processes that underpin them.

5.1.1 Aims and Objectives

Using the set of highly robust and non-redundant radiomic features identified in Chapter 4, we aimed to investigate the predictive value of these radiomic features for the multi-class classification of culprit, non-culprit and asymptomatic carotid arteries in symptomatic and asymptomatic patients respectively. We compared the predictive performance of radiomic features with carotid calcification and PET imaging information. In addition, we aimed to assess the biological relevance of these radiomic features by correlating the radiomic features with functional PET data and histological findings from carotid endarterectomy specimens.

The specific objectives for this chapter are as follows:

Objectives:

- To assess the predictive ability of robust and non-redundant radiomic features in multi-class classification (asymptomatic versus culprit versus non-culprit carotid arteries) with feature-based machine learning.
- To investigate the predictive ability of a) highly robust features only, b) all radiomic features and c) poorly robust features only to determine whether using robust features improves model performance and generalisability.
- To assess the impact of integrating radiomic features with clinical information (PET data and carotid calcium score) on predictive ability.
- To interpret machine learning outputs with Shapley Additive Explanations (SHAP).
- To assess radiomic feature correlations with PET imaging data.
- To assess radiomic feature correlations with %CD68+ as determined by immunohistochemistry.

5.1.2 Hypotheses

- There is greater predictive accuracy in multi-class classification than binary classification.
- Using only robust radiomic features as predictors leads to better generalisation than using *all* radiomic features or using *poorly robust* radiomic features as predictors.
- Integrating radiomic features with PET data and carotid calcium can improve discriminatory ability.
- We can gain some insight into the importance of different predictors for multi-class classification decisions according to the machine learning models.
- Radiomic features will relate to biological markers of inflammation as determined by PET imaging and immunohistochemistry.

5.2 Materials and Methods

5.2.1 Carotid CT Angiography Scans and PyRadiomics CTA Features

In this chapter, the radiomic features extracted from carotid CT angiography data in Chapter 3 and identified as robust in Chapter 4 with PyRadiomics were further investigated. Please refer to Chapter 2, section 2.2.1 for details regarding carotid CT angiography data acquisition, to Chapter 3, section 3.2.2 for details of PyRadiomics radiomic feature extraction and to Chapter 4, section 4.2.2 for details regarding the identification of robust radiomic features.

5.2.2 Carotid PET information

The CTA images analysed in this study had associated PET imaging acquired from the PET-CT Department of Addenbrooke's Hospital on a GE Discovery 690 (GE Medical Systems Ltd, Hatfield, UK) with 64-slice computed tomography. Further details of PET imaging methodology and scanning parameters can be found in the original publications of VISION[152], CHAI[153] and ICARUSS[151] which used established methods for vascular PET imaging. In each respective study, different PET tracers were investigated. These are summarised in Table 5.72, along with the targets of the respective radiotracers.

Table 5.72 PET Tracer Name and Target

Tracer Name	Chemical Name	Target
FDG	^{18}F-fluorodeoxyglucose	Allows quantification of ^{18}F-2-deoxy-D-glucose uptake within the artery wall; associated with increased metabolic rate
DOTATATE	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid - [Tyr3]octreotate	Binds to somatostatin receptor type 2 (SST2); reports on macrophage-related inflammation
NaF	^{18}F-sodium fluoride	Binds to hydroxyapatite, allowing for quantification of microcalcification within atherosclerotic plaques.
FMISO	^{18}F-fluoromisonidazole	Highly specific for hypoxia; accumulates nonlinearly as the partial pressure of oxygen falls below 10 mmHg.

As different PET tracers and different combinations of the tracers were investigated in the various studies, not all of the carotid arteries analysed with radiomics have information for all of the PET tracers listed in Table 5.72.

Of the 132 carotid arteries in the radiomics dataset (82 from symptomatic patients [41 culprit; 41 non-culprit carotid arteries] and 50 carotid arteries from asymptomatic patients):

- 120 carotid arteries had corresponding FDG PET data.
- 56 carotid arteries had DOTATATE PET data.
- 44 carotid arteries had NaF PET data.
- 20 carotid arteries had FMISO PET data.

In VISION, ICARUSS and CHAI, PET radiotracer uptake was measured by first calculating standardised uptake values (SUV). SUV was calculated by taking the target tissue concentration of the radiotracer and correcting this for injected dose and body weight. The tissue-to-background ratio (TBR) can be calculated by dividing vessel SUV by blood pool venous tracer activity and has become the standard measure of tracer activity in vascular PET studies[204]. If TBR was calculated from the maximum SUV value within an ROI, the resulting value is called the TBR_{max}.

Previously, in multi-slice analysis (as described in Chapter 3, section 3.2.2.4), 14 slices of the carotid CTA scans (resampled to 3 mm slice thickness) were analysed in each volume-of-interest to derive 93 radiomic features. This approach had been taken to match with how their corresponding PET scans were analysed in the original studies[151–153], whereby PET images were co-registered with CT and resampled to 3mm slice thickness with multi-slice analysis involving ROIs drawn on 3 slices below and 10 slices above the carotid bifurcation.

To summarise the radiotracer activity amongst all these slices in the multi-slice approach, the mean of the TBR_{max} value was calculated by JT (for VISION studies), NE (for ICARUSS studies) and FJ (for CHAI studies). This value was derived by averaging the TBR_{max} over all the analysed axial slices in the multi-slice approach, see Figure 5.59.

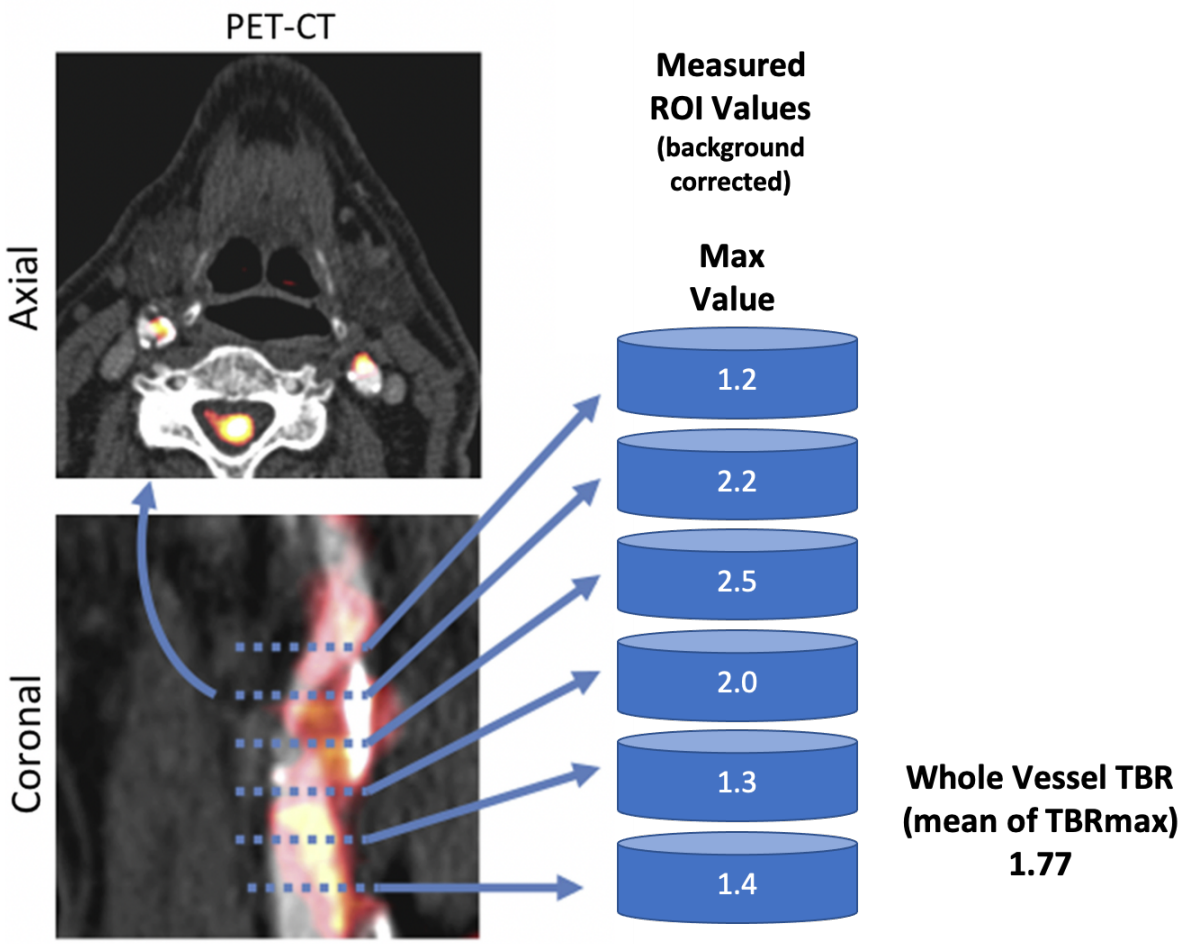


Figure 5.59 Schema of TBR_{max} calculation

Regions of interest (ROI) are drawn around the artery (in axial orientation) to provide maximum standardized uptake values (SUV) for each ROI. This is repeated along the length of the vessel for 14 slices of the carotid artery. The background corrected maximum SUVs = TBR_{max} per slice. These values are averaged to produce the mean of the TBR_{max} for multi-slice analysis. In this schema, the example has been simplified to involve 6 slices only, using FDG-PET/CT. This figure has been adapted from Tawakol et al. 2013[205] *and Evans, 2017*[206].

5.2.3 Multi-class Feature-based Machine Learning

5.2.3.1 Data pre-processing and predictors

Multi-class classification involved distinguishing between culprit, non-culprit and asymptomatic carotid arteries.

Features from several settings were investigated as predictors:

1. Multi-slice approach without resegmentation (original image)
    a. Using the 14 highly robust and non-redundant radiomic features identified in Chapter 4 as predictors
    b. Using all 93 radiomic features derived in multi-slice analysis without resegmentation as predictors
    c. Using the 10 poorly robust radiomic features as predictors
2. Multi-slice approach with resegmentation
    a. Using the 10 highly robust and non-redundant radiomic features as predictors
    b. Using all 93 radiomic features derived in multi-slice analysis with resegmentation as predictors
    c. Using the 2 poorly robust radiomic features as predictors

These different approaches were investigated to assess the hypothesis that using robust radiomic features would lead to better generalisable performance than non-robust radiomic features.

The performance of these radiomic features was also compared with the predictive ability of carotid calcification alone and FDG PET mean TBR_{max} alone. FDG PET was chosen rather than the other tracers as there were more FDG scans available. Finally, an integrated model that combined the highly robust and non-redundant radiomic features with FDG PET mean TBR_{max} was also investigated. The training set for the machine learning algorithms involved features derived from the carotid CTA scans of the 120 carotid arteries that had corresponding FDG PET values. As in Chapter 4, each predictor (individual radiomic feature) was standardised to have a mean of zero and a variance of one prior to machine learning classification. This was important to correct for how the different radiomic features varied in their value ranges.

5.2.3.2 Machine Learning Classifiers and Model Performance Evaluation

The machine learning classifiers assessed, configurations and random state used were consistent with those previously described in Chapter 4, section 4.2.8.3, to include a decision tree, random forest, LASSO regression, Elastic Net regression, a neural network and XGBoost, implemented with sklearn version 0.23.1.

To estimate the out-of-sample performance of the machine learning classifiers (i.e. expected generalisability on unseen data) (1) stratified five-fold cross validation was performed (reporting the mean accuracy and AUC across the five folds) and (2) bootstrapping (1000 repetitions). Please note that in order to calculate the AUC for this multi-class classification task (as opposed to a binary classification task), we employed the One-vs-Rest (also known as One-vs-All) strategy. This strategy computes the average of the AUC scores for each class against all other classes.

The bootstrapping method is a resampling technique used to estimate summary statistics such as mean accuracy, as well as to provide 95% confidence intervals by repeatedly sampling a dataset with replacement. Bootstrapping was described further in Chapter 1, section 1.4.4.

The performance estimates for accuracy, which was calculated by cross validation and bootstrapping, was compared with the performance on an external validation set that comprised 12 carotid arteries (the 12 carotid arteries without FDG PET values from the original 132 carotid artery dataset). In this way, the out-of-sample performance as determined by cross-validation and by bootstrapping could be compared to actual performance on external data. Consequently, with 132 carotid arteries used for training and 12 carotid arteries used for external validation, this represented a 90:10 train-test split (90% of the data set was used for training, 10% used for testing) which is a commonly used ratio for dividing the data[207].

In addition, the sensitivity and specificity for each carotid artery class was calculated using the One-vs-All strategy for multi-class classifications. Table 5.73 provides the formulas for calculating the respective evaluation metrics.

Table 5.73 Formulas for sensitivity and specificity

Evaluation metric	Equation
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
TP, true positives; FN, false negatives; TN, true negatives; FP, false positives	

In multi-class classification when there are three classes (i.e. asymptomatic [Asx], culprit [CC] and non-culprit [NC] carotid arteries), using the One-vs-All strategy, TP, FN, TN and FP are calculated as follows[208]:

- "TP of Asx" is all Asx instances (ground truth) that were correctly classified as asymptomatic.
- "TN of Asx" is all non-Asx instances (i.e. CC or NC) that were not classified as asymptomatic.
- "FP of Asx" is all non-Asx instances that were classified as asymptomatic by the model.
- "FN of Asx" is all Asx instances that were not classified as asymptomatic.

Replace Asx with CC or NC respectively to obtain the necessary terms in the sensitivity and specificity equations.

5.2.3.3 Interpretability

To better interpret the classification decisions of the highest performing machine learning classifiers and gain further insights into the behaviour of the individual predictors, the SHAP (SHapley Additive exPlanations) approach was used[209]. The SHAP explanation method illustrates the contribution of each predictor in terms of feature importance for a classifier's predictions and whether each predictor contributes positively or negatively to the carotid artery classification (i.e. asymptomatic, culprit or non-culprit) in terms of SHAP values.

SHAP values are computed based on Shapley values from 'coalitional game theory'[210], in which the features are the players in the game and the reward is a prediction of 1 e.g. towards predicting the 'culprit' carotid class. Therefore, a positive SHAP value indicates driving the prediction towards 1 whilst a negative SHAP value drives the prediction towards 0, i.e. the

other carotid classes. SHAP values can be used to explain the output of any ML model, therefore, the SHAP method is a model-agnostic prediction explainer[211]. The open source code for the SHAP approach can be found here: https://github.com/slundberg/shap.

5.2.4 Histology and Immunohistochemistry of Carotid Plaque Specimens

5.2.4.1 Dataset of culprit carotid plaque specimens

To be a useful quantitative imaging biomarker, radiomic features must not only possess useful discriminatory ability, but should relate in some rational way to the pathophysiology underlying the medical image[212]. Carotid specimens were only available from the VISION study and involved specimens that corresponded to the culprit carotid plaque from 7 symptomatic patients who underwent carotid endarterectomy. Slices of these carotid specimens were stained with 3,3'-diaminobenzidine (DAB) and counterstained with haematoxylin. Immunostaining of the carotid sections was performed using the pan-macrophage marker CD68 to allow for visualisation of macrophage infiltration into the carotid plaque. Further details of the histological preparation of these carotid specimens and of the immunostaining procedure can be found in the original thesis of Dr Tarkin (JT, 2017), section 2.8[213].

5.2.4.2 Quantification of carotid plaque macrophage infiltration

Images of these sections were taken using ImagePro Insight (Media Cybernetics, Silver Spring, USA) by Parvesh Konda (PNK, 4th year Cambridge medical student during his student selected component). A quantitative method was used to measure the percentage area stained representing macrophages (%CD68+) in the images using Fiji (version 1.52p)[214]. The method followed that of Jenson (2013)[215], with the addition of a colour deconvolution step as recommended by Shen (2019)[216].

The quantification process was as follows:

Step 1: Colour deconvolution as implemented by Fiji was applied to the images captured by ImagePro. This automatic feature in Fiji allows for separation of the coloured image into 3 separate channels according to histological stain-specific red-green-blue (RGB) absorption[217].

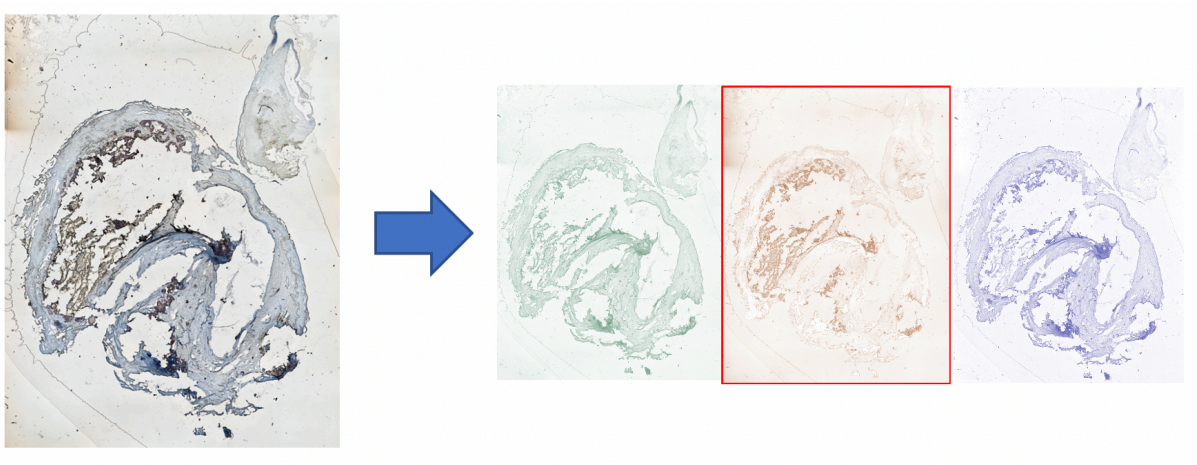


Figure 5.60 Colour deconvolution step.

The image representing macrophage staining was selected. The selected image is outlined with a red border.

Step 2: Thresholding was applied to the deconvolved image in order to highlight the areas of macrophage staining only. Thresholding was also applied to the original coloured image to highlight the whole tissue. The threshold levels applied were agreed by consensus between PNK and EL.

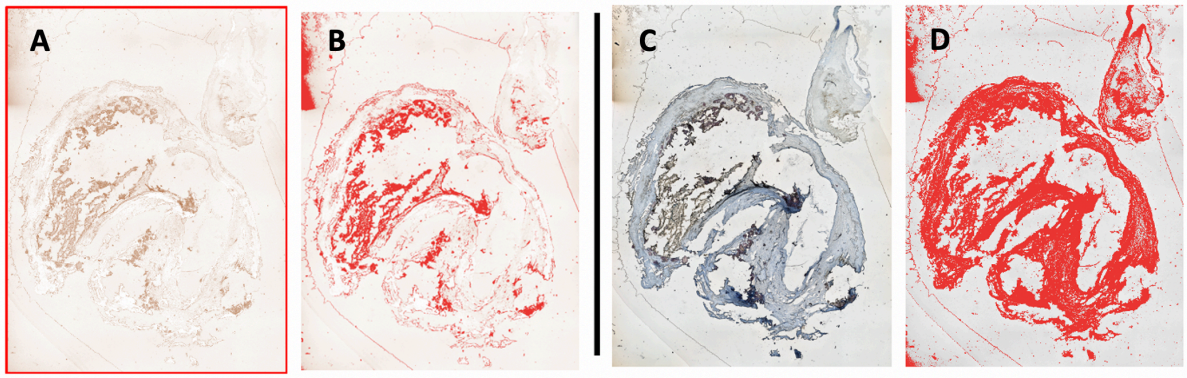


Figure 5.61 Thresholding the deconvolved image and the original image

A) Deconvolved image, B) Deconvolved image after thresholding for macrophages, C) Original image, D) Original image after thresholding for stained tissue.

Step 3: Regions-of-interest (ROIs) were independently drawn by EPVL and PNK to outline the highlighted areas on the deconvolved image after thresholding for macrophages and for the original image after thresholding for stained tissue.

Step 4: We subsequently used Fiji's analyse function to automatically measure the area of the stained tissue, resulting in a measure of the area stained for macrophages and a measure of the total area of the whole tissue. The %CD68+ could then be quantified by dividing the area stained for macrophages by the total area and multiplying by 100. The final %CD68+ per slice was obtained by averaging the measures by EPVL and PNK.

5.2.4.3 Radiomic Feature Correlations with Immunohistochemistry

There were 7 symptomatic patients that had carotid CTA radiomics and histology with immunohistochemical staining of their culprit carotid plaque for the presence of macrophages (CD68) after carotid endarterectomy. The histology slides were oriented to the clinical images by JT to identify the portion of carotid artery from which the carotid plaque section corresponded to, i.e. (1) carotid bifurcation, (2) above the carotid bifurcation or (3) below the carotid bifurcation. The %CD68+ in the histology slides were subsequently correlated with the CTA radiomic features of the corresponding axial CTA slice.

5.2.5 Statistical Analysis

All continuous data are given as means $\pm$ standard deviation or median (IQR) as appropriate. A two-tailed *p-value*<0.05 was considered to indicate statistical significance. Data was tested for normality using the Shapiro-Wilk test. To assess the correlation of the radiomic features with PET mean of TBR_{max} or with %CD68+ from carotid specimens, the non-parametric Spearman's rank correlation test was performed.

5.3 Results

5.3.1 Multi-class Feature-based Machine Learning Classification

5.3.1.1 Highly Robust Radiomic Features

Table 5.74 provides a list of the highly robust radiomic features identified in Chapter 4, section 4.3.4 for multi-slice analysis: (1) original image (no prior resegmentation/image normalisation) and (2) with resegmentation. These radiomic feature sets were used as predictors for the multi-class classification of culprit, non-culprit and asymptomatic carotid arteries.

Table 5.74 Multi-slice analysis radiomic feature sets with excellent robustness

Multi-slice (B-spline interpolation, Bin width=25)	
Original (n=14)	**Resegmentation (n=10)**
15. First Order: 90th Percentile 16. First Order: Root Mean Squared 17. First Order: Maximum 18. First Order: Robust Mean Absolute Deviation 19. GLCM: Cluster Shade 20. GLCM: Cluster Tendency 21. GLCM: Inverse Variance 22. GLCM: Joint Average 23. GLDM: Large Dependence Emphasis 24. GLDM: Large Dependence High Grey Level Emphasis 25. GLSZM: Small Area Emphasis 26. GLSZM: Size Zone NonUniformity 27. NGTDM: Busyness 28. NGTDM: Complexity	11. First Order: Mean Absolute Deviation 12. GLCM: Difference Variance 13. GLCM: Joint Energy 14. GLDM: Dependence Variance 15. GLDM: Large Dependence High Grey Level Emphasis 16. GLRLM: Long Run High Grey Level Emphasis 17. GLRLM: Grey Level NonUniformity Normalised 18. GLSZM: Grey Level NonUniformity 19. GLSZM: Large Area High Grey Level Emphasis 20. GLSZM: Size Zone NonUniformity

The training data consisted of the radiomic features detailed in Table 5.74 that were derived from carotid CTA scans of 120 carotid arteries. The performance of (1) the highly robust radiomic features alone, (2) the combination of FDG PET mean of TBR_{max} and radiomic features and of (3) FDG PET mean of TBR_{max} alone were evaluated.

5.3.1.1.1 Multi-slice: Original Image

Using the five-fold cross validation scheme, the LASSO classifier and the Elastic Net classifier had the highest AUC, with both reaching an AUC of 0.81 when using radiomic features with excellent robustness only as predictors. This was better performance than using FDG PET mean of TBR_{max} alone (see Table 5.76), which achieved its highest AUC of 0.59 with a neural network classifier. Combining the PET information with the radiomic features led to similar performance with using radiomic features alone, see Table 5.75 for details of the individual classifier performances.

Table 5.75 Multi-slice (Original image): Five-fold Cross Validation

Classifier	Predictors					
	Calcium + Radiomic Features		PET + Radiomic Features		Radiomic Features only	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Decision Tree	0.54 (0.05)	0.66 (0.03)	0.52 (0.01)	0.64 (0.07)	0.56 (0.08)	0.67 (0.06)
Random Forest	0.61 (0.14)	0.77 (0.08)	0.59 (0.15)	0.77 (0.09)	0.58 (0.13)	0.79 (0.09)
LASSO	0.63 (0.07)	0.82 (0.08)	0.62 (0.09)	0.80 (0.08)	0.65 (0.06)	0.81 (0.08)
Elastic Net	0.63 (0.08)	0.81 (0.08)	0.60 (0.07)	0.79 (0.08)	0.64 (0.08)	0.81 (0.07)
Neural Network	0.58 (0.05)	0.77 (0.05)	0.54 (0.03)	0.74 (0.04)	0.58 (0.06)	0.77 (0.06)
XGBoost	0.55 (0.12)	0.73 (0.09)	0.59 (0.13)	0.73 (0.09)	0.55 (0.13)	0.75 (0.11)

Mean accuracy and AUC reported with standard deviation in brackets. The radiomic features used here are those that were identified to have excellent robustness in the multi-slice approach (original image i.e. no image pre-processing) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

Using carotid calcium alone for multi-class classification was better than chance, with its highest AUC reaching 0.69 with a neural network, see Table 5.76. However, this was comparatively worse than using radiomic features only as predictors, Table 5.75. Depending on the classifier used, combining carotid calcium scores with the highly robust radiomic features could slightly decrease or slightly improve the performance – overall calcium had little additional impact.

Table 5.76 Five-fold Cross Validation Performance of PET and Carotid Calcium only

Classifier	**Predictors**			
	FDG PET mean of TBR_{max} only		**Carotid Calcium Only**	
	Accuracy	**AUC**	**Accuracy**	**AUC**
Decision Tree	0.33 (0.09)	0.50 (0.07)	0.54 (0.07)	0.60 (0.08)
Random Forest	0.33 (0.09)	0.50 (0.07)	0.53 (0.09)	0.66 (0.09)
LASSO	0.38 (0.06)	0.56 (0.05)	0.45 (0.07)	0.68 (0.06)
Elastic Net	0.38 (0.06)	0.56 (0.05)	0.45 (0.07)	0.68 (0.05)
Neural Network	0.40 (0.09)	0.59 (0.07)	0.45 (0.06)	0.69 (0.05)
XGBoost	0.36 (0.11)	0.51 (0.06)	0.49 (0.13)	0.61 (0.09)

Mean accuracy and AUC reported with standard deviation in brackets.

Estimates for the out-of-sample performance of the machine learning classifiers investigated in Table 5.75 were also calculated using a bootstrapping procedure which provided a mean accuracy and AUC as well as 95% confidence intervals for these performance statistics. The estimates obtained from five-fold cross-validation and bootstrapping were similar, with LASSO and Elastic Net having the highest AUC when using radiomic features only as predictors.

Table 5.77 Multi-slice (Original image): Bootstrapping

Classifier	**Predictors**					
	Calcium + Radiomic Features		**PET + Radiomic Features**		**Radiomic Features only**	
	Accuracy	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
Decision Tree	0.53 CI: 0.40 to 0.65	0.65 CI: 0.56 to 0.74	0.52 CI: 0.40 to 0.65	0.64 CI: 0.55 to 0.74	0.53 CI: 0.40 to 0.64	0.65 CI: 0.56 to 0.73
Random Forest	0.56 CI: 0.45 to 0.67	0.76 CI: 0.68 to 0.83	0.55 CI: 0.44 to 0.66	0.75 CI: 0.67 to 0.82	0.55 CI: 0.43 to 0.67	0.75 CI: 0.67 to 0.83
LASSO	0.60 CI: 0.47 to 0.71	0.78 CI: 0.69 to 0.86	0.59 CI: 0.46 to 0.71	0.78 CI: 0.69 to 0.86	0.61 CI: 0.49 to 0.73	0.79 CI: 0.70 to 0.86
Elastic Net	0.60 CI: 0.48 to 0.72	0.78 CI: 0.69 to 0.86	0.59 CI: 0.47 to 0.71	0.78 CI: 0.69 to 0.85	0.61 CI: 0.49 to 0.72	0.79 CI: 0.71 to 0.86
Neural Network	0.58 CI: 0.47 to 0.69	0.76 CI: 0.67 to 0.84	0.56 CI: 0.45 to 0.68	0.75 CI: 0.66 to 0.82	0.58 CI: 0.46 to 0.70	0.76 CI: 0.68 to 0.84
XGBoost	0.55 CI: 0.44 to 0.66	0.74 CI: 0.65 to 0.81	0.55 CI: 0.43 to 0.65	0.73 CI: 0.64 to 0.81	0.55 CI: 0.43 to 0.67	0.74 CI: 0.65 to 0.82

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping. The radiomic features used here are those that were identified to have excellent robustness in the multi-slice approach (original image i.e. no image pre-processing) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

The performance of calcium alone as a predictor as determined by bootstrapping was comparable to that of FDG PET mean of TBR_{max} alone, both of which performed similarly to chance alone, see Table 5.78. Although broadly similar estimates arose from cross-validation and bootstrapping, bootstrapping provided slightly less optimistic estimates of out-of-sample performance.

Table 5.78 Bootstrapping Performance of PET and Calcium Predictors

Classifier	**Predictors**			
	FDG PET mean of TBR_{max} only		**Carotid Calcium Only**	
	Accuracy	**AUC**	**Accuracy**	**AUC**
Decision Tree	0.33 CI: 0.23 to 0.44	0.51 CI: 0.42 to 0.59	0.36 CI: 0.23 to 0.48	0.51 CI: 0.41 to 0.62
Random Forest	0.33 CI: 0.23 to 0.44	0.50 CI: 0.41 to 0.61	0.35 CI: 0.24 to 0.47	0.52 CI: 0.41 to 0.65
LASSO	0.35 CI: 0.24 to 0.45	0.57 CI: 0.48 to 0.65	0.32 CI: 0.19 to 0.43	0.54 CI: 0.41 to 0.64
Elastic Net	0.35 CI: 0.24 to 0.45	0.57 CI: 0.48 to 0.65	0.32 CI: 0.19 to 0.44	0.54 CI: 0.43 to 0.3
Neural Network	0.35 CI: 0.25 to 0.45	0.56 CI: 0.46 to 0.64	0.33 CI: 0.21 to 0.45	0.53 CI: 0.42 to 0.63
XGBoost	0.35 CI: 0.23 to 0.46	0.51 CI: 0.42 to 0.60	0.36 CI: 0.24 to 0.48	0.52 CI: 0.41. to 0.64

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping.

Since the LASSO logistic regression classifier using radiomic features with excellent robustness only as the predictors had the highest AUC and accuracy as determined by cross validation and bootstrapping, it was chosen as the model to further evaluate in terms of interpretability with SHAP analysis and for assessment of its performance on the external validation set.

Figure 5.62 demonstrates a SHAP feature importance plot which ranks the predictors according to how much they contribute to a model's predictions according to their mean absolute SHAP value. It provides the global importance of the radiomic features for predictions based on the original training dataset (120 carotid arteries). Overall, First Order: Root Mean Squared was the most influential feature, particularly for differentiating culprit carotid arteries. For asymptomatic carotid arteries, the most important feature was GLDM: Large Dependence High Grey Level Emphasis and for non-culprit carotid arteries, the most important feature was First Order: 90th Percentile.

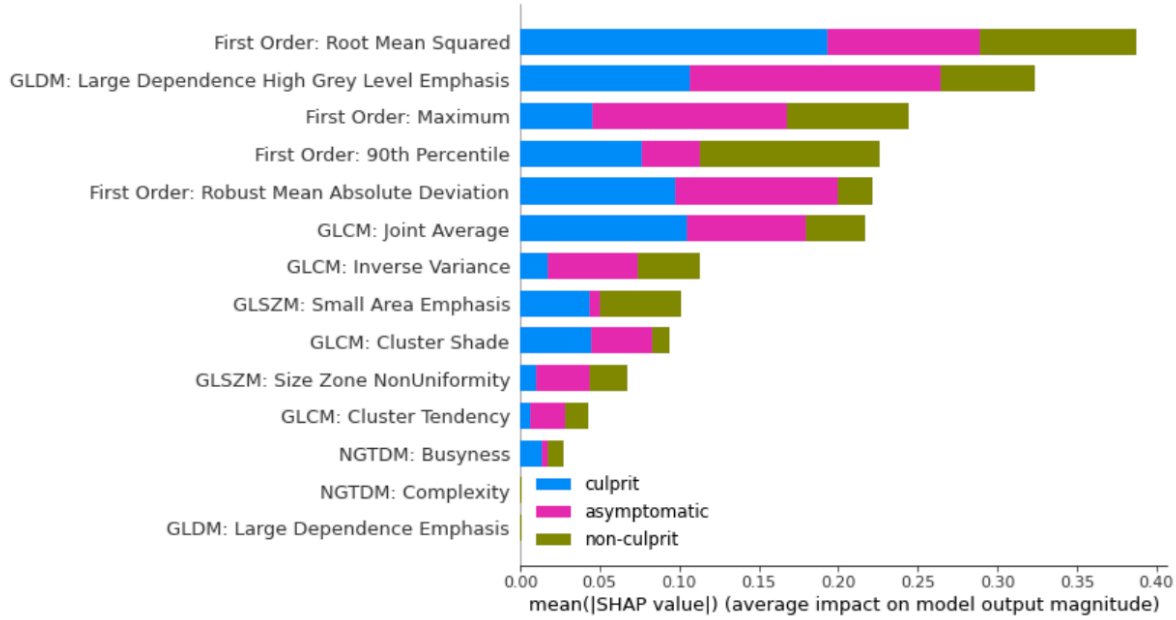


Figure 5.62 Multi-slice (Original) Highly Robust Radiomic Features (LASSO) SHAP Feature Importance Plot for the Training Dataset

This SHAP feature importance plot provides the global importance of the radiomic features with excellent robustness in multi-slice analysis (original image) in the LASSO classifier's predictions on the training dataset in identifying culprit, non-culprit and asymptomatic carotid arteries. Features with larger absolute SHAP values are more important than the others. The features are sorted in order of decreasing importance on the y-axis. The features at the top of the plot contribute more to the model than the ones at the bottom.

To gain a deeper understanding of how the individual features contribute (i.e. positively or negatively) towards predicting a certain carotid artery class, SHAP summary value plots were generated (Figure 5.63, Figure 5.64, Figure 5.65). These plots provide information, not only about feature importance, but also feature effects. Each dot in the plot represents a SHAP value for a feature and a data instance. The colour of the dot represents the value of the feature which ranges from low (blue) to high (red). As with the SHAP feature importance plot, the features are ordered by decreasing importance.

Figure 5.63 provides the SHAP summary value plot for the asymptomatic carotid artery class and therefore demonstrates how the individual highly robust multi-slice (original image) radiomic features contribute towards or away from predicting the asymptomatic class. In general, the higher the value of GLDM: Large Dependence High Grey Level Emphasis, the more likely the model would predict 'asymptomatic' based on the training dataset.

Figure 5.64 provides the plot for the culprit carotid artery class which demonstrates that high values of First Order: Root Mean Squared drove the prediction away from 'culprit' carotid arteries and towards asymptomatic/non-culprit carotid arteries. Therefore, the lower the value of First Order: Root Mean Squared, the more likely the model would predict the 'culprit' carotid artery class based on the training dataset. Lower values of GLDM: Large Dependence High Grey Level Emphasis were also associated with the culprit carotid artery class.

Figure 5.65 provides the plot for the non-culprit carotid artery class which was associated with lower values of First Order: 90^{th} Percentile, but higher values of First Order: Root Mean Squared. A higher First Order: Maximum made it more likely that the model would predict 'non-culprit' carotid artery based on the training dataset.

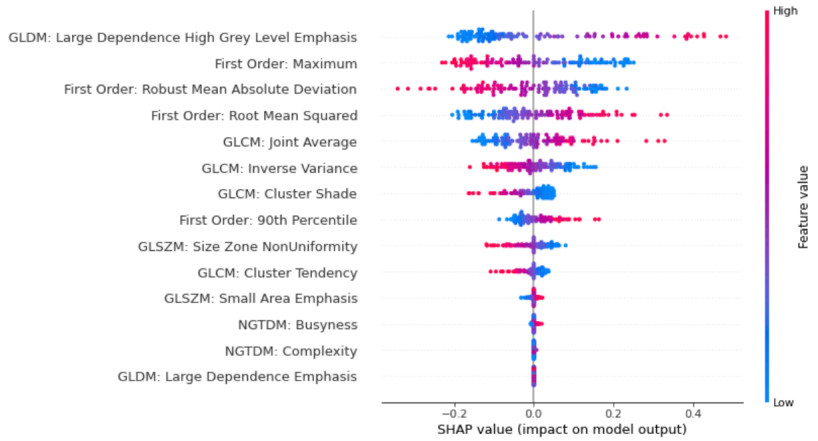


Figure 5.63 Multi-slice (Original) Highly Robust Radiomic Features (LASSO) SHAP Summary Value Plot for the Training Data: Asymptomatic Carotid Artery Class

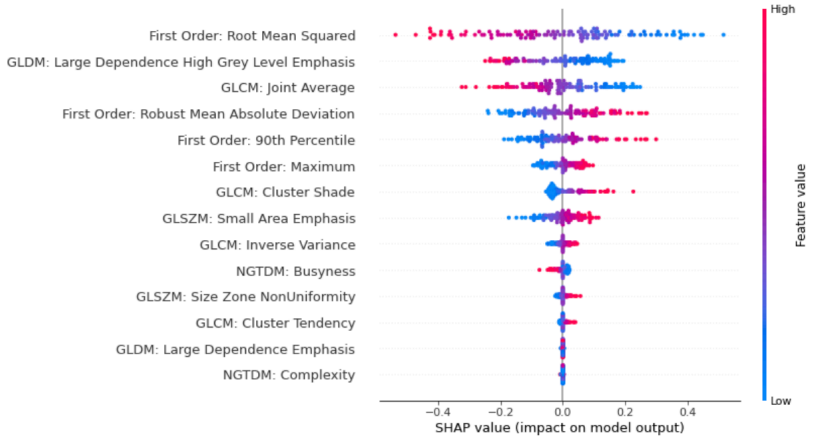


Figure 5.64 Multi-slice (Original) Highly Robust Radiomic Features (LASSO) SHAP Summary Value Plot for the Training Data: Culprit Carotid Artery Class

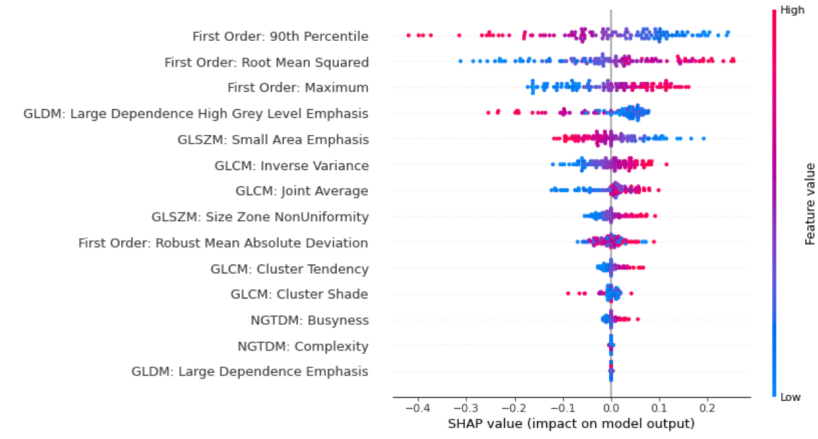


Figure 5.65 Multi-slice (Original) Highly Robust Radiomic Features (LASSO) SHAP Summary Value Plot for the Training Data: Non-culprit Carotid Artery Class

The SHAP summary value plots provide information about feature importance and whether each predictor contributes positively or negatively to the target variable i.e. carotid artery type (asymptomatic, culprit or non-culprit). The features are ranked in descending order according to feature importance, the horizontal location of the dots demonstrates whether the effect of that feature value drives towards (more rightward) or away (more leftward) from the carotid artery class and the colour represents the feature value with high values being red and low values being blue.

The external validation set consisted of 12 carotid arteries (10 asymptomatic, 1 culprit and 1 non-culprit carotid artery). This was similar to a carotid type distribution in the general population, but different to the distribution of the training set. The training set had an equal distribution of asymptomatic (n=40), culprit (n=40) and non-culprit (n=40) carotid arteries. Table 5.79 presents a confusion matrix of the classifier's predictions versus the ground truth status of the carotid arteries. The classifier identified all asymptomatic carotid arteries correctly but did not correctly identify the culprit or the non-culprit carotid artery. Overall, the LASSO classifier achieved an external testing accuracy of 83.3% and an AUC of 0.86. The sensitivity and specificity respectively for each carotid artery class was as follows: Asx (100%, 50%), CC (0%, 91%) and NC (0%, 100%).

Table 5.79 Multi-slice (Original Image): External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**10**	0	0
CC	1	**0**	0
NC	0	1	**0**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the LASSO classifier are highlighted in red. Predictors: multi-slice analysis (original image) radiomic features with excellent robustness only.

5.3.1.1.2 Multi-slice: Resegmentation

The performance of highly robust radiomic features in multi-slice analysis with resegmentation (see Table 5.80) was largely similar to those in multi-slice analysis without resegmentation (original image, see Table 5.75). When using the radiomic features only as predictors, the neural network achieved the highest AUC of 0.80, although it was closely followed by the Elastic Net, LASSO and Random Forest classifiers. This was greater than the performance achieved by FDG PET mean of TBR_{max} alone or carotid calcification alone (see Table 5.76) and better than the combination of PET with radiomic features (see Table 5.80). The highest AUC achieved (0.84) in this setting was the combination of highly robust radiomic features (multi-slice with resegmentation) and carotid calcification, using the neural network model.

Table 5.80 Multi-Slice (Resegmentation): Five-fold Cross Validation

Classifier	Predictors					
	Calcium + Radiomic Features		PET + Radiomic Features		Radiomic Features only	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Decision Tree	0.55 (0.05)	0.66 (0.04)	0.57 (0.09)	0.68 (0.07)	0.59 (0.05)	0.69 (0.04)
Random Forest	0.66 (0.07)	0.83 (0.06)	0.63 (0.08)	0.77 (0.08)	0.63 (0.06)	0.77 (0.08)
LASSO	0.67 (0.05)	0.83 (0.06)	0.58 (0.07)	0.76 (0.09)	0.63 (0.08)	0.77 (0.08)
Elastic Net	0.70 (0.06)	0.82 (0.06)	0.60 (0.08)	0.77 (0.08)	0.63 (0.09)	0.78 (0.07)
Neural Network	0.68 (0.04)	0.84 (0.04)	0.55 (0.04)	0.78 (0.06)	0.62 (0.04)	0.80 (0.06)
XGBoost	0.61 (0.04)	0.81 (0.04)	0.59 (0.03)	0.76 (0.07)	0.61 (0.06)	0.75 (0.07)

Mean accuracy and AUC reported with standard deviation in brackets. The radiomic features used here are those that were identified to have excellent robustness in the multi-slice approach (with resegmentation) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

The AUC estimates as determined by bootstrapping resulted in similar estimates to five-fold cross validation. However, cross-validation sometimes provided slightly more optimistic estimates than bootstrapping. As such, the remainder of this study will estimate out-of-sample performance with bootstrapping only. The neural network demonstrated equivalent performance compared with Elastic Net, LASSO and Random Forest classifiers, see Table 5.81.

Table 5.81 Multi-slice (Resegmentation): Bootstrapping

Classifier	**Predictors**					
	Calcium + Radiomic Features		**PET + Radiomic Features**		**Radiomic Features only**	
	Accuracy	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
Decision Tree	0.58 CI: 0.44 to 0.69	0.69 CI: 0.58 to 0.77	0.52 CI: 0139 to 0.64	0.64 CI: 0.55 to 0.73	0.52 CI: 0.38 to 0.64	0.64 CI: 0.54 to 0.74
Random Forest	0.61 CI: 0.50 to 0.71	0.80 CI: 0.72 to 0.87	0.59 CI: 0.47 to 0.70	0.76 CI: 0.68 to 0.83	0.59 CI: 0.47 to 0.69	0.76 CI: 0.68 to 0.83
LASSO	0.61 CI: 0.49 to 0.74	0.79 CI: 0.69 to 0.86	0.57 CI: 0.44 to 0.68	0.75 CI: 0.66 to 0.83	0.57 CI: 0.45 to 0.70	0.76 CI: 0.66 to 0.84
Elastic Net	0.61 CI: 0.48 to 0.73	0.79 CI: 0.69 to 0.86	0.56 CI: 0.43 to 0.67	0.74 CI: 0.65 to 0.82	0.57 CI: 0.45 to 0.69	0.76 CI: 0.67 to 0.84
Neural Network	0.62 CI: 0.50 to 0.73	0.79 CI: 0.71 to 0.87	0.54 CI: 0.42 to 0.66	0.73 CI: 0.64 to 0.82	0.57 CI: 0.45 to 0.67	0.76 CI: 0.66 to 0.84
XGBoost	0.61 CI: 0.49 to 0.71	0.79 CI: 0.71 to 0.86	0.57 CI: 0.45 to 0.68	0.74 CI: 0.66 to 0.82	0.58 CI: 0.47 to 0.69	0.75 CI: 0.67 to 0.82

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping. The radiomic features used here are those that were identified to have excellent robustness in the multi-slice approach (with resegmentation) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

Multi-slice analysis with resegmentation: radiomic features with excellent robustness only SHAP analysis

When using highly robust radiomic features only (multi-slice analysis with resegmentation) as predictors, the neural network had the highest AUC as determined by cross-validation and was chosen for SHAP analysis. Figure 5.66 demonstrates that overall, the most important feature for carotid artery discrimination was GLCM: Difference Variance, particularly for predicting the culprit carotid arteries. GLDM: Dependence Variance played a larger role in predicting the asymptomatic carotids, as well as for predicting non-culprit carotid arteries.

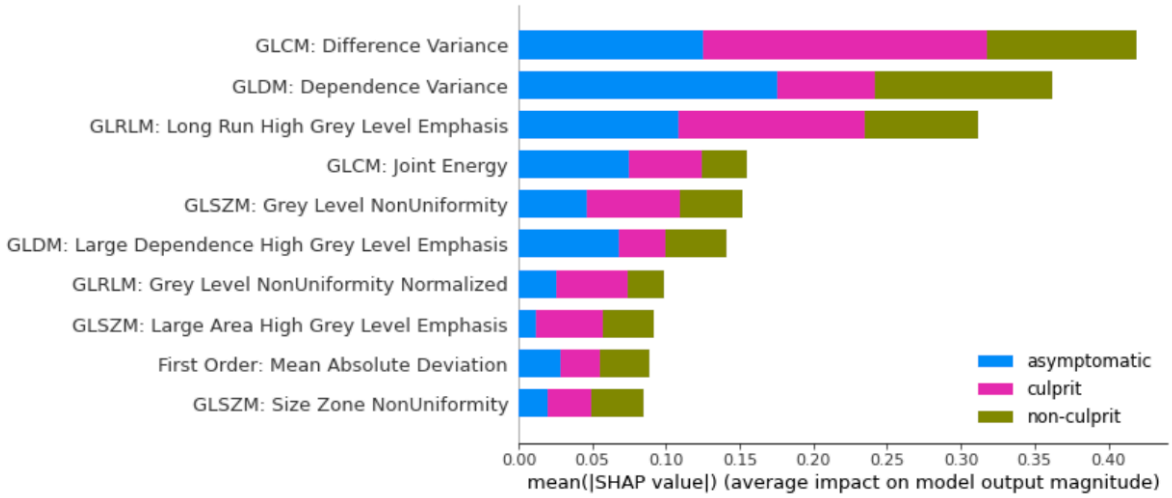


Figure 5.66 Multi-slice (Resegmentation) Highly Robust Radiomic Features (Neural Network) SHAP Feature Importance Plot for the Training Dataset

This SHAP feature importance plot provides the global importance of the radiomic features with excellent robustness in multi-slice analysis (resegmentation) in the neural network classifier's predictions on the training dataset in identifying culprit, non-culprit and asymptomatic carotid arteries.

Figure 5.67 demonstrates that lower values of GLDM: Dependence Variance, but higher values of GLCM: Difference Variance increased the likelihood of predicting the asymptomatic class in the training dataset. The asymptomatic class was also associated with higher values of GLRLM: Long Run High Grey Level Emphasis and GLCM: Joint Energy. Figure 5.68 shows that culprit carotid arteries were associated with lower values of GLDM: Difference Variance and lower values of GLRLM: Long Run High Grey Level Emphasis. Non-culprit carotids were associated with higher values of GLCM: Difference Variance and GLDM: Dependence Variance, see Figure 5.69.

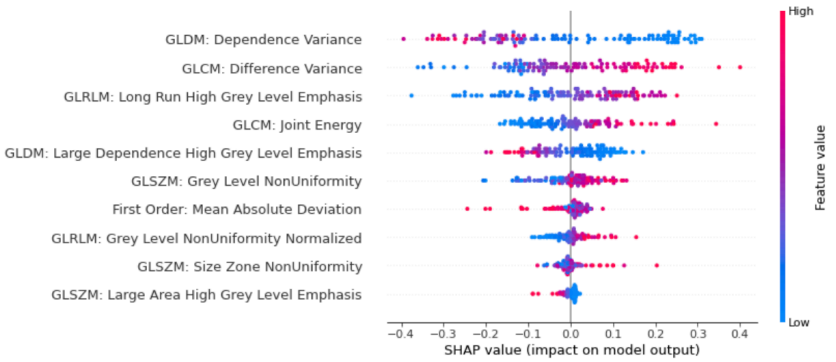


Figure 5.67 Multi-slice (Resegmentation) Highly Robust Radiomic Features (Neural Network) SHAP Summary Value Plot for the Training Dataset: Asymptomatic Carotid Artery Class

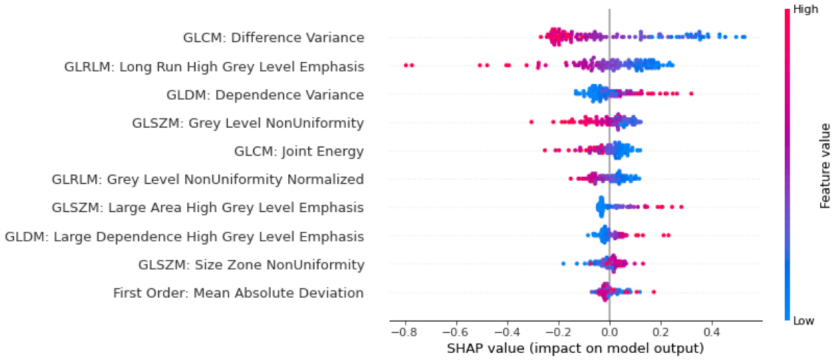


Figure 5.68 Multi-slice (Resegmentation) Highly Robust Radiomic Features (Neural Network) SHAP Summary Value Plot for the Training Dataset: Culprit Carotid Artery Class

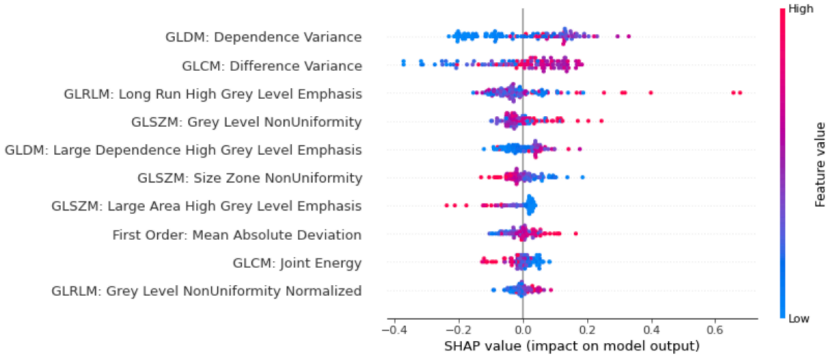


Figure 5.69 Multi-slice (Resegmentation) Highly Robust Radiomic Features (Neural Network) SHAP Summary Value Plot for the Training Dataset: Non-culprit Carotid Artery Class

The SHAP summary value plots provide information about feature importance and whether each predictor contributes positively or negatively to the target variable i.e. carotid artery type (asymptomatic, culprit or non-culprit). The features are ranked in descending order according to feature importance, the horizontal location of the dots demonstrates whether the effect of that feature value drives towards (more rightward) or away (more leftward) from the carotid artery class and the colour represents the feature value with high values being red and low values being blue.

Multi-slice analysis with resegmentation: radiomic features with excellent robustness only (neural network) external validation performance

The neural network achieved an external testing accuracy of 75.0% with an AUC of 0.61. Although this was better than chance, using radiomic features following resegmentation for multi-class classification led to a lesser performance compared with multi-slice (original image) radiomic features. The classifier identified the majority of asymptomatic carotid arteries correctly, however, confused one of the asymptomatic carotid arteries for a culprit carotid artery. The culprit and the non-culprit carotid artery predictions were incorrect, see Table 5.82. The sensitivity and specificity respectively for each carotid artery class was as follows: Asx (90%, 50%), CC (0%, 82%) and NC (0%, 100%).

Table 5.82 Multi-slice (Resegmentation): External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**9**	1	0
CC	1	**0**	0
NC	0	1	**0**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the neural network classifier are highlighted in red. Predictors: multi-slice analysis (resegmentation) radiomic features with excellent robustness only.

Multi-slice analysis with resegmentation: radiomic features with excellent robustness and carotid calcification SHAP analysis

The neural network classifier using the highly robust radiomic features and carotid calcium led to the highest predictive performance with internal validation. SHAP analysis demonstrated that (1) GLDM: Dependence Variance, (2) Calcification and (3) GLCM: Difference Variance were the most important features in the model's predictions on the training data, see Figure 5.70.

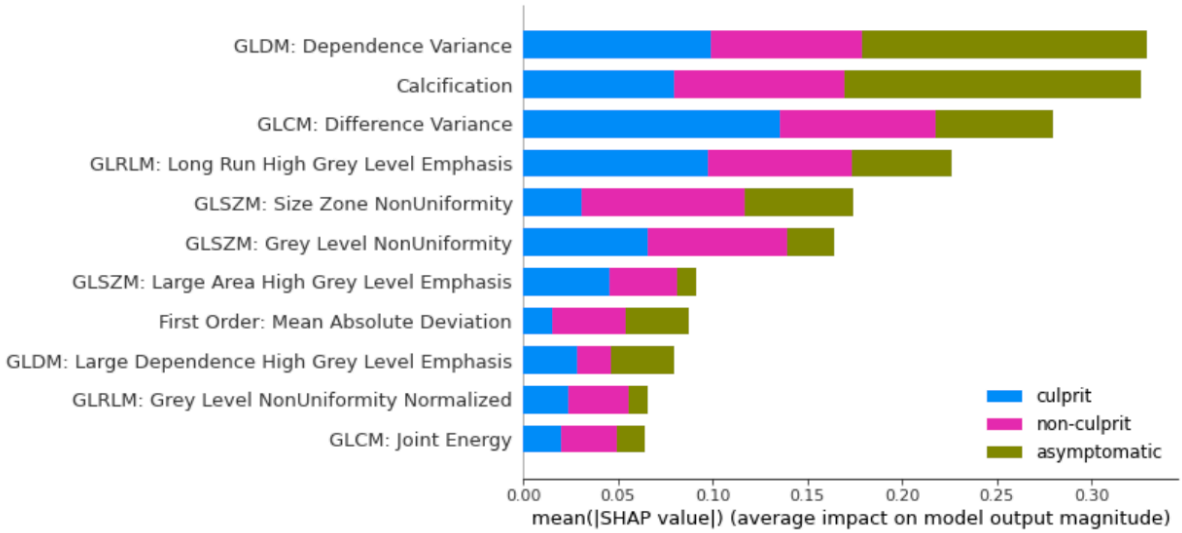


Figure 5.70 Multi-slice (Resegmentation) Highly Robust Radiomic Features and Carotid Calcification (Neural Network) SHAP Feature Importance Plot for the Training Dataset

Figure 5.71 demonstrates that lower values of calcification were associated with predicting the asymptomatic carotid artery class, whilst higher values of calcification were associated with predicting the non-culprit carotid artery class (see Figure 5.73). Asymptomatic carotid arteries were associated with lower values of GLDM: Dependence Variance, whilst culprit carotid arteries were associated with higher values (see Figure 5.72).

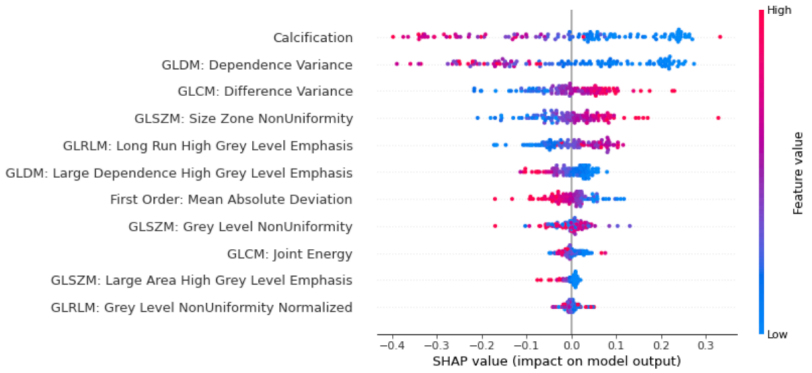


Figure 5.71 Multi-slice (Resegmentation) Highly Robust Radiomic Features and Carotid Calcification (Neural Network) SHAP Summary Value Plot for the Training Dataset: Asymptomatic Carotid Artery Class

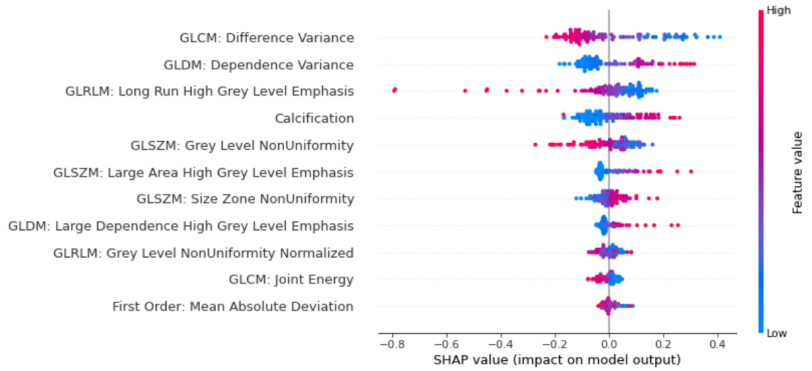


Figure 5.72 Multi-slice (Resegmentation) Highly Robust Radiomic Features and Carotid Calcification (Neural Network) SHAP Summary Value Plot for the Training Dataset: Culprit Carotid Artery Class

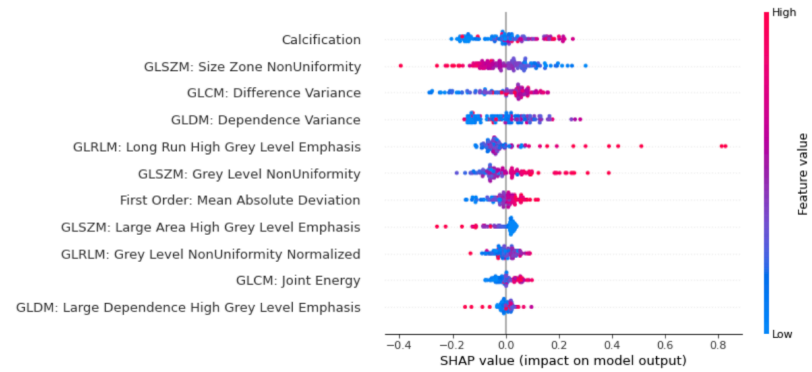


Figure 5.73 Multi-slice (Resegmentation) Highly Robust Radiomic Features and Carotid Calcification (Neural Network) SHAP Summary Value Plot for the Training Dataset: Non-culprit Carotid Artery Class

The SHAP summary value plots provide information about feature importance and whether each predictor contributes positively or negatively to the target variable i.e. carotid artery type (asymptomatic, culprit or non-culprit). The features are ranked in descending order according to feature importance, the horizontal location of the dots demonstrates whether the effect of that feature value drives towards (more rightward) or away (more leftward) from the carotid artery class and the colour represents the feature value with high values being red and low values being blue.

Multi-slice analysis with resegmentation: radiomic features with excellent robustness and carotid calcification (neural network) external validation performance

Upon testing of the neural network classifier on the external validation dataset, a test accuracy of 75.0% and an AUC of 0.81 was achieved. Although there was a slight decrease in the correct classifications of asymptomatic carotid arteries, the non-culprit carotid artery was correctly identified. However, the culprit carotid artery was still predicted as an asymptomatic carotid artery. The sensitivity and specificity respectively of each carotid artery class was as follows: Asx (80%, 50%), CC (0%, 100%) and NC (100%, 91%).

Table 5.83 Multi-slice (Resegmentation): External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	8	1	1
CC	1	**0**	0
NC	0	0	**1**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the neural network classifier are highlighted in red. Predictors: multi-slice analysis (resegmentation) radiomic features with excellent robustness and carotid calcification.

5.3.1.2 All Radiomic Features

5.3.1.2.1 Multi-slice: Original Image

All 93 radiomic features used in this analysis are detailed in Chapter 3, Table 3.2.

Table 5.84 Multi-slice (Original image): Bootstrapping

Classifier	Predictors					
	Calcium + Radiomic Features		PET + Radiomic Features		Radiomic Features only	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Decision Tree	0.56 CI: 0.42 to 0.69	0.67 CI: 0.57 to 0.77	0.55 CI: 0.42 to 0.67	0.67 CI: 0.57. to 0.76	0.56 CI: 0.43 to 0.69	0.67 CI: 0.58 to 0.77
Random Forest	0.60 CI: 0.47 to 0.71	0.79 CI: 0.70 to 0.87	0.59 CI: 0.47 to 0.71	0.79 CI: 0.70 to 0.86	0.59 CI: 0.47 to 0.71	0.79 CI: 0.70 to 0.86
LASSO	0.62 CI: 0.50 to 0.74	0.80 CI: 0.72 to 0.87	0.62 CI: 0.50 to 0.73	0.80 CI: 0.72 to 0.87	0.62 CI: 0.51 to 0.74	0.80 CI: 0.72 to 0.87
Elastic Net	0.62 CI: 0.51 to 0.73	0.80 CI: 0.72 to 0.87	0.62 CI: 0.50 to 0.74	0.80 CI: 0.71 to 0.87	0.62 CI: 0.51 to 0.74	0.80 CI: 0.72 to 0.87
Neural Network	0.61 CI: 0.50 to 0.72	0.79 CI: 0.70 to 0.86	0.60 CI: 0.48. to 0.71	0.78 CI: 0.68 to 0.86	0.61 CI: 0.49 to 0.72	0.78 CI: 0.69 to 0.86
XGBoost	0.60 CI: 0.47 to 0.72	0.79 CI: 0.70 to 0.86	0.60 CI: 0.48 to 0.72	0.79 CI: 0.69 to 0.86	0.60 CI: 0.48 to 0.72	0.79 CI: 0.70 to 0.86

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping. The radiomic features used here are all 93 radiomic features extracted in the multi-slice approach (original image) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

The LASSO and Elastic Net classifiers achieved similar performances using radiomic features only, with or without PET or carotid calcium information. For testing on the external dataset and SHAP analysis, the LASSO model was chosen according to Occam's razor which encourages using the simpler model when they achieve the same performance.

Multi-slice analysis (original): all radiomic features SHAP analysis

The top 20 radiomic feature predictors in multi-slice analysis (original image) are displayed in Figure 5.74. First Order: Median was the most influential feature overall for the LASSO classifier predictions on the training dataset, playing a particularly large role in predicting culprit carotid artery status.

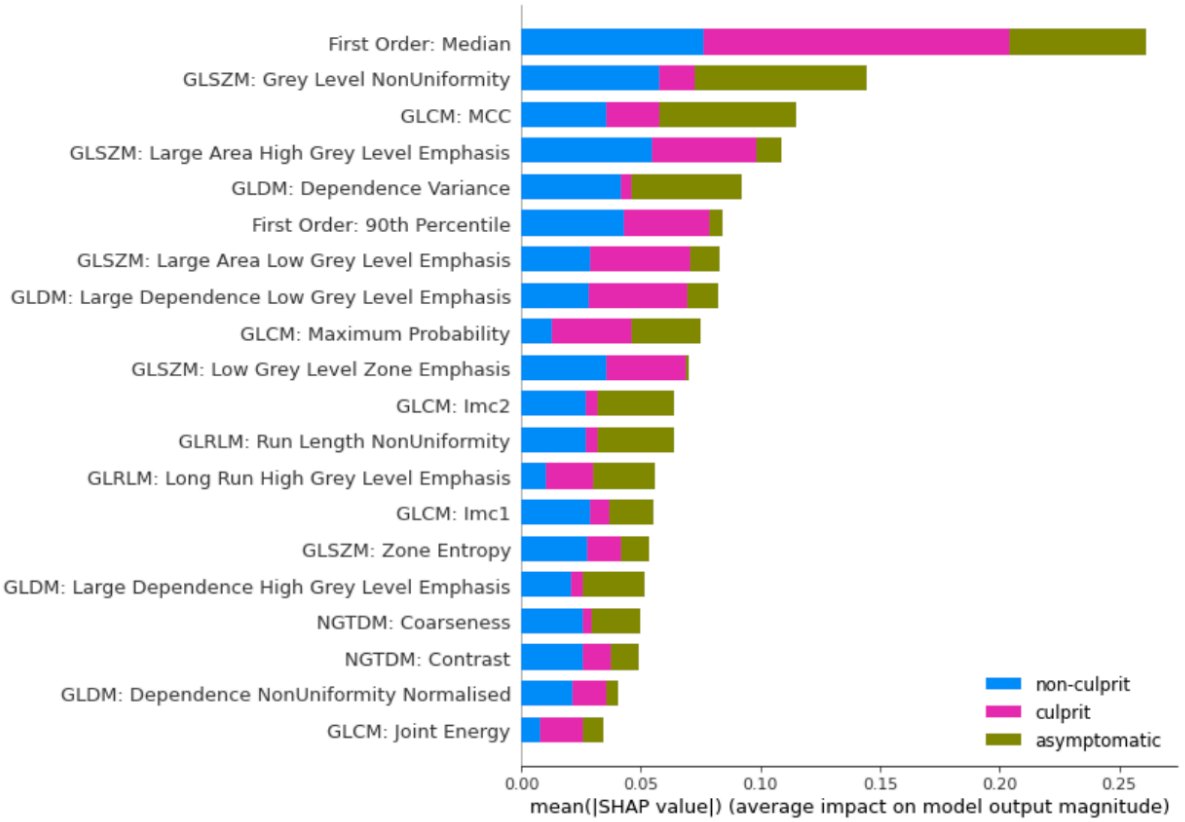


Figure 5.74 Multi-slice (original) all radiomic features (LASSO) SHAP Feature Importance Plot for the Training Dataset

Higher values of GLSZM: Grey Level NonUniformity, First Order: Median and of GLDM: Dependence Variance were associated with predicting the asymptomatic class. Lower values of First Order: Median and of GLSZM: Large Area High Grey Level Emphasis were associated with predicting the culprit carotid class. Culprit carotid arteries were also associated with higher values of GLSZM: Large Area Low Grey Level Emphasis and GLDM: Large Dependence Low Grey Level Emphasis. On the other hand, non-culprit carotid arteries were associated with a higher First Order: Median and GLSZM: Large Area High Grey Level Emphasis, but lower values of GLSZM: Grey Level NonUniformity.

Multi-slice analysis (original): all radiomic features (LASSO) external validation performance

Using the LASSO classifier and all radiomic features as predictors, a testing accuracy of 83.3% and an AUC of 0.73 was achieved. Similar to how the model using only the highly robust multi-slice (original) radiomic features performed, all asymptomatic carotid arteries were correctly identified, however the culprit and the non-culprit carotid artery were not. The sensitivity and specificity respectively for each carotid artery class was as follows: Asx (100%, 50%), CC (0%, 91%), NC (0%, 100%).

Table 5.85 Multi-slice (Original): All Radiomic Features External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**10**	0	0
CC	1	**0**	0
NC	0	1	**0**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the LASSO classifier are highlighted in red. Predictors: multi-slice analysis (original) all 93 radiomic features.

Multi-slice analysis (original): all radiomic features and carotid calcium SHAP analysis and external validation performance

Following the incorporation of carotid calcium along with all radiomic features, the importance of the top 20 radiomic features (see Figure 5.75) were largely the same with the model that incorporated only the radiomic features as predictors, as was the performance on the external validation set. The LASSO classifier achieved an accuracy of 83.3% and an AUC of 0.73. The SHAP summary plots for the respective classes were also largely similar to the radiomic features only model. Carotid calcification was not in the top 20 predictors for any carotid artery class in the presence of radiomic features derived from the multi-slice analysis (original image) setting.

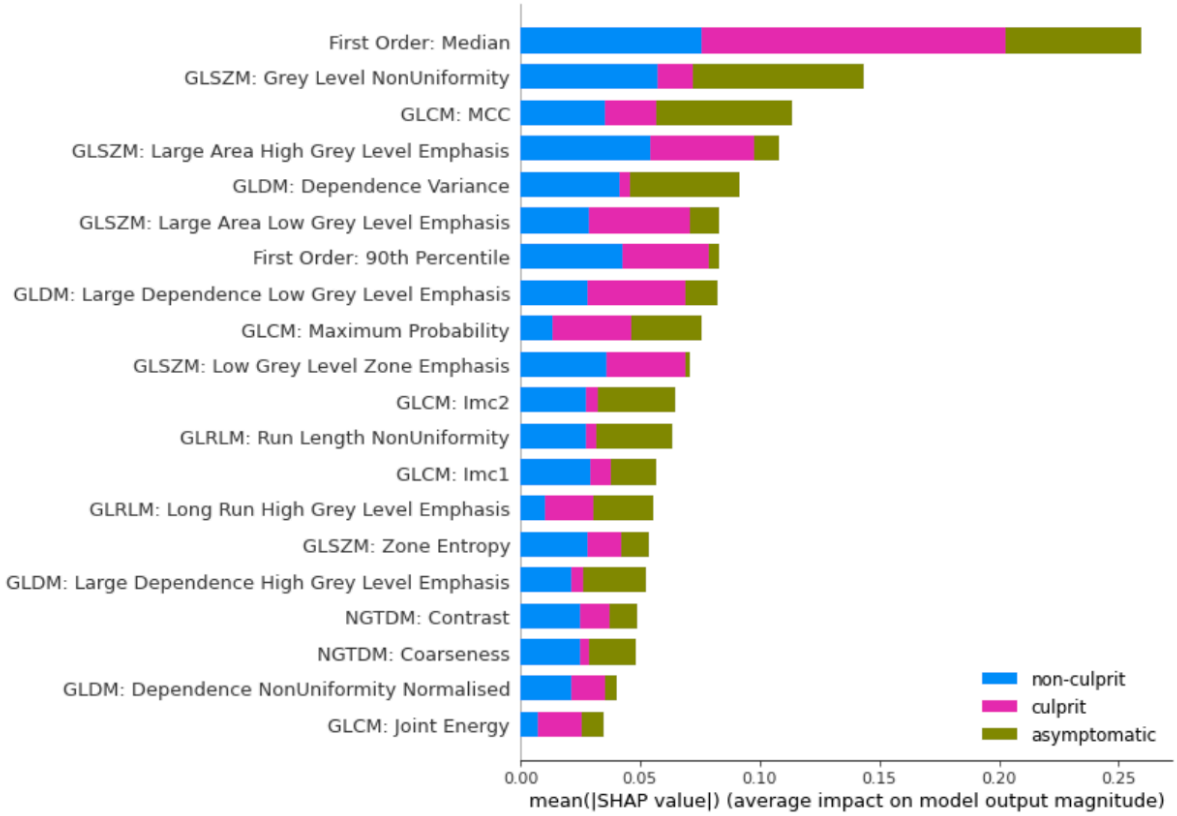


Figure 5.75 Multi-slice (original) all radiomic features and carotid calcification (LASSO) SHAP Feature Importance Plot for the External Validation Set

5.3.1.2.2 Multi-slice: Resegmentation

The random forest classifier with all radiomic features (multi-slice with resegmentation) and carotid calcium led to the highest predictive performance according to bootstrapping in the multi-slice (resegmentation) setting, see Table 5.86. However, the performance was similar to using multi-slice (original) radiomic features, see Table 5.84.

Table 5.86 Multi-slice (Resegmentation): All Radiomic Features - Bootstrapping

Classifier	**Predictors**					
	Calcium + Radiomic Features		**PET + Radiomic Features**		**Radiomic Features only**	
	Accuracy	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
Decision Tree	0.56 CI: 0.42 to 0.70	0.68 CI: 0.57 to 0.77	0.54 CI: 0.40 to 0.67	0.66 CI: 0.55 to 0.75	0.53 CI: 0.40 to 0.66	0.65 CI: 0.55 to 0.75
Random Forest	0.63 CI: 0.51 to 0.75	0.81 CI: 0.73 to 0.87	0.62 CI: 0.51 to 0.73	0.79 CI: 0.72 to 0.86	0.62 CI: 0.51 to 0.73	0.79 CI: 0.72 to 0.86
LASSO	0.61 CI: 0.48 to 0.72	0.78 CI: 0.70 to 0.86	0.59 CI: 0.46 to 0.70	0.77 CI: 0.68 to 0.85	0.59 CI: 0.46 to 0.70	0.77 CI: 0.68 to 0.85
Elastic Net	0.61 CI: 0.48 to 0.72	0.78 CI: 0.70 to 0.86	0.58 CI: 0.45 to 0.70	0.77 CI: 0.68 to 0.85	0.58 CI: 0.45 to 0.71	0.77 CI: 0.68 to 0.85
Neural Network	0.61 CI: 0.49 to 0.72	0.78 CI: 0.69 to 0.86	0.58 CI: 0.46 to 0.69	0.77 CI: 0.67 to 0.85	0.59 CI: 0.47 to 0.71	0.77 CI: 0.68 to 0.85
XGBoost	0.62 CI: 0.49 to 0.73	0.80 CI: 0.71 to 0.87	0.59 CI: 0.47 to 0.70	0.77 CI: 0.69 to 0.85	0.58 CI: 0.46 to 0.70	0.77 CI: 0.68 to 0.85

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping. The radiomic features used here are all 93 radiomic features extracted in the multi-slice approach (resegmentation) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

Multi-slice analysis (resegmentation): all radiomic features with carotid calcification SHAP analysis

The top 20 predictors in multi-slice analysis (resegmentation) are displayed in Figure 5.76 when all radiomic features and carotid calcification were used as predictors. The most important predictors overall for the training dataset were (1) NGTDM: Complexity and (2) carotid calcification.

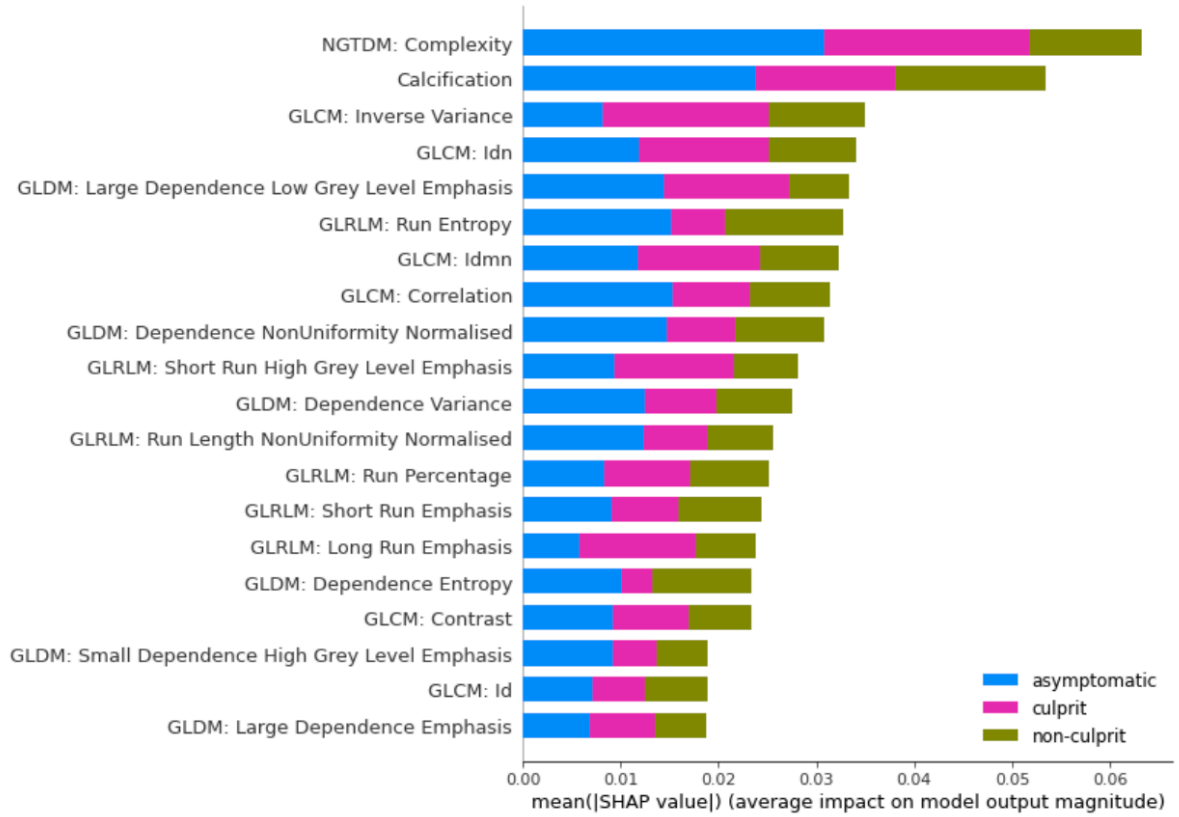


Figure 5.76 Multi-slice (Resegmentation) All Radiomic Features and Carotid Calcification (Random Forest) SHAP Feature Importance Plot for the Training Dataset

Asymptomatic carotid arteries were associated with higher values for NGTDM: Complexity and. For GLDM: Dependence NonUniformity Normalised. However, it was associated with lower values of carotid calcification and GLCM: Correlation. Culprit carotid arteries were associated with lower values of NGTDM: Complexity, but higher values of GLCM: Inverse Variance, Calcification and GLDM: Idn. Non-culprit carotid arteries were associated with moderate to higher values of calcification, GLRLM: Run Entropy and NGTDM: Complexity.

Multi-slice analysis (resegmentation): all radiomic features and carotid calcification (Random Forest) external validation performance

The random forest classifier reached a test accuracy of 83.3% and an AUC of 0.82 on the external validation dataset. This was a slightly better performance than when only highly robust radiomic features were used with carotid calcification (multi-slice with resegmentation), see Table 5.83. Although the culprit carotid artery was still misclassified as an asymptomatic carotid artery, more of the asymptomatic carotid arteries were correctly classified when all radiomic features were used compared with highly robust features only. The sensitivity and specificity respectively for each carotid artery class was as follows: Asx (90%, 50%), CC (0%, 100%), NC (100%, 91%).

Table 5.87 Multi-slice (Resegmentation): External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**9**	0	1
CC	1	**0**	0
NC	0	0	**1**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the random forest classifier are highlighted in red. Predictors: multi-slice analysis (resegmentation) all radiomic features and carotid calcification.

5.3.1.3 Poorly Robust Radiomic Features

The poorly robust radiomic features in multi-slice analysis (original image) were (1) First Order: 10th Percentile, (2) First Order: Minimum, (3) GLDM: Low Grey Level Emphasis, (4) GLDM: Small Dependence Low Grey Level Emphasis, (5) GLRLM: Long Run Low Grey Level Emphasis, (6) GLRLM: Low Grey Level Run Emphasis, (7) GLRLM: Run Length NonUniformity, (8) GLRLM: Short Run Low Grey Level Emphasis, (9) GLSZM: Low Grey Level Zone Emphasis and (10) GLSZM: Small Area Low Grey Level Emphasis.

The poorly robust radiomic features in multi-slice analysis (resegmentation) were (1) First Order: Minimum and (2) First Order: Range.

5.3.1.3.1 Multi-slice: Original Image

The neural network using the poorly robust radiomic features with carotid calcification had the highest performance according to bootstrapping, see Table 5.88. The use of poorly robust radiomic features alone still led to higher performance than PET information or carotid calcification alone.

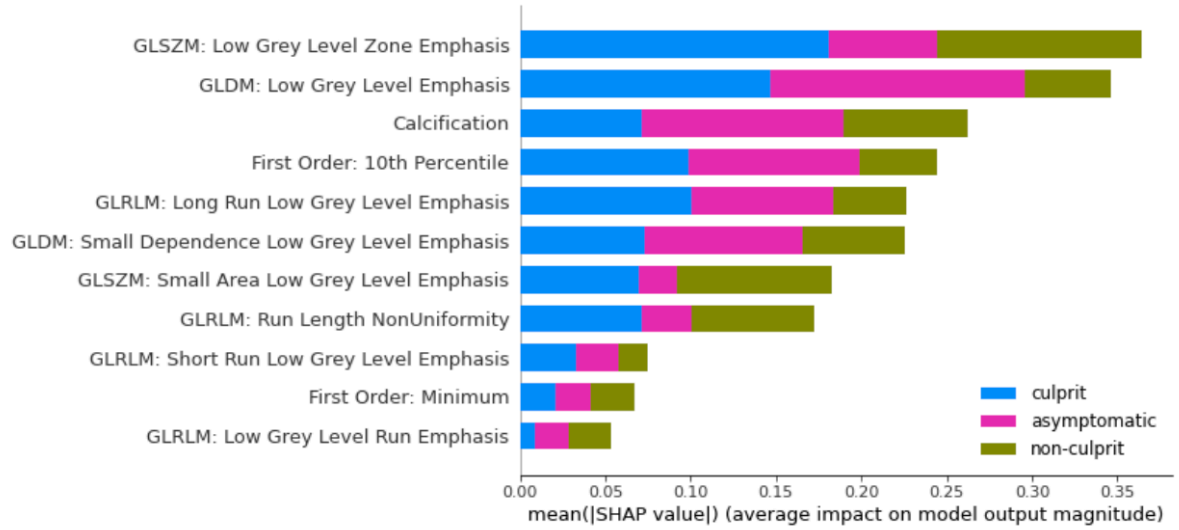


Figure 5.77 Multi-slice (Original) Poorly Robust Radiomic Features and Carotid Calcification (Neural Network) SHAP Feature Importance Plot for the Training Dataset

SHAP analysis revealed that GLSZM: Low Grey Level Zone Emphasis and. GLDM: Low Grey Level Emphasis, followed by carotid calcification were the most important predictors overall for the training dataset, see Figure 5.77.

Table 5.88 Multi-slice (Original image): Poorly Robust Features - Bootstrapping

Classifier	Predictors					
	Calcium + Radiomic Features		PET + Radiomic Features		Radiomic Features only	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Decision Tree	0.50 CI: 0.36 to 0.63	0.63 CI: 0.52 to 0.73	0.44 CI: 0.32 to 0.58	0.59 CI: 0.50 to 0.69	0.46 CI: 0.33 to 0.59	0.60 CI: 0.51 to 0.70
Random Forest	0.56 CI: 0.43 to 0.69	0.75 CI: 0.66 to 0.83	0.50 CI: 0.38 to 0.63	0.70 CI: 0.61 to 0.78	0.51 CI: 0.38 to 0.63	0.71 CI: 0.61 to 0.79
LASSO	0.56 CI: 0.43 to 0.67	0.74 CI: 0.65 to 0.81	0.53 CI: 0.41 to 0.65	0.73 CI: 0.64 to 0.81	0.54 CI: 0.42 to 0.65	0.74 CI: 0.66 to 0.81
Elastic Net	0.56 CI: 0.44 to 0.69	0.74 CI: 0.66 to 0.82	0.53 CI: 0.41 to 0.64	0.73 CI: 0.64 to 0.81	0.54 CI: 0.42 to 0.67	0.74 CI: 0.65 to 0.82
Neural Network	0.57 CI: 0.45 to 0.68	0.75 CI: 0.66 to 0.83	0.51 CI: 0.39 to 0.63	0.70 CI: 0.61 to 0.79	0.54 CI: 0.41 to 0.66	0.73 CI: 0.65 to 0.81
XGBoost	0.54 CI: 0.42 to 0.66	0.73 CI: 0.64 to 0.82	0.48 CI: 0.36 to 0.60	0.67 CI: 0.58 to 0.77	0.50 CI: 0.38 to 0.62	0.69 CI: 0.59 to 0.78

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping. The radiomic features used here are the poorly robust radiomic features extracted in the multi-slice approach (original image) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

SHAP summary value plots (not shown) demonstrated that asymptomatic carotid arteries were associated with lower values of GLDM: Low Grey Level Emphasis and carotid calcification, but higher values of First Order: 10^{th} Percentile and GLDM: Small Dependence Low Grey Level Emphasis. Culprit carotid arteries were associated with lower values of GLSZM: Low Grey Level Zone Emphasis and First Order: 10^{th} Percentile, but higher values of GLDM: Low Grey Level Emphasis and GLRLM: Long Run Grey Level Emphasis. Non-culprit carotid arteries were associated with higher values of GLSZM: Low Grey Level Zone Emphasis, Calcification and GLRLM: Run Length NonUniformity, but lower values of GLSZM: Small Area Low Grey Level Emphasis.

The neural network classifier achieved a test accuracy of 75% and an AUC of 0.88 on the external validation dataset. The majority of asymptomatic carotid arteries were correctly classified. The culprit carotid was predicted as the non-culprit carotid and this was the other way around for the non-culprit carotid artery. The sensitivity and specificity respectively for each class was as follows: Asx (90%, 100%), CC (0%, 91%) and NC (0%, 82%).

Table 5.89 Multi-slice (Original): External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**9**	0	1
CC	0	**0**	1
NC	0	1	**0**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the neural network classifier are highlighted in red. Predictors: multi-slice analysis (original) poorly robust radiomic features and carotid calcification.

5.3.1.3.2 Multi-slice: Resegmentation

Table 5.90 Multi-slice (Resegmentation): Poorly Robust Features - Bootstrapping

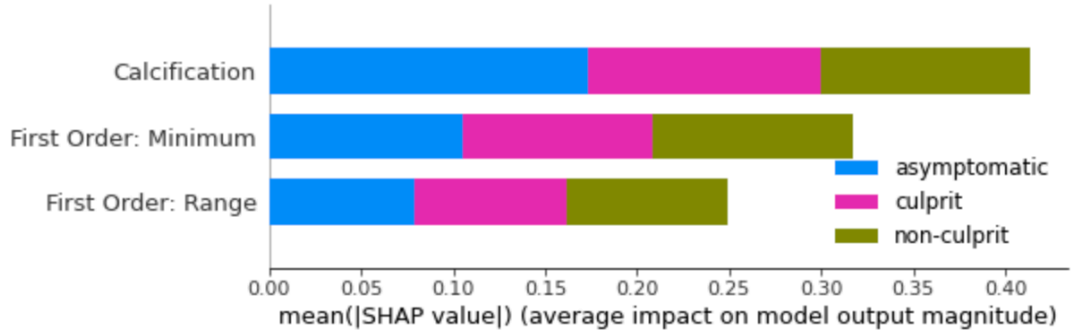
Classifier	Predictors					
	Calcium + Radiomic Features		PET + Radiomic Features		Radiomic Features only	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Decision Tree	0.43 CI: 0.31 to. 0.55	0.58 CI: 0.49 to 0.66	0.34 CI: 0.24 to 0.46	0.51 CI: 0.43 to 0.60	0.32 CI: 0.22 to 0.44	0.50 CI: 0.42 to 0.58
Random Forest	0.43 CI: 0.31 to 0.55	0.61 CI: 0.50 to 0.70	0.35 CI: 0.24 to 0.45	0.52 CI: 0.43 to 0.61	0.34 CI: 0.22 to 0.43	0.50 CI: 0.41 to 0.59
LASSO	0.37 CI: 0.23 to. 0.50	0.60 CI: 0.46 to 0.69	0.34 CI: 0.23 to. 0.44	0.55 CI: 0.46 to 0.64	0.29 CI: 0.20 to 0.38	0.48 CI: 0.40 to 0.56
Elastic Net	0.38 CI: 0.23 to 0.50	0.60 CI: 0.47 to 0.69	0.34 CI: 0.23 to 0.44	0.55 CI: 0.46 to 0.64	0.29 CI: 0.20 to 0.38	0.48 CI: 0.40 to 0.57
Neural Network	0.40 CI: 0.27 to. 0.50	0.59 CI: 0.50 to 0.67	0.32 CI: 0.22 to 0.43	0.53 CI: 0.43 to 0.61	0.28 CI: 0.19 to 0.37	0.47 CI: 0.38 to 0.55
XGBoost	0.42 CI: 0.31 to 0.53	0.61 CI: 0.51 to 0.71	0.36 CI: 0.25 to 0.47	0.53 CI 0.44 to 0.62	0.33 CI: 0.23 to 0.43	0.51 CI: 0.41 to 0.60

Mean accuracy and AUC reported with CI (95% confidence intervals) as determined by bootstrapping. The radiomic features used here are the poorly robust radiomic features extracted in the multi-slice approach (resegmentation) setting. The PET predictor was the FDG mean of TBR_{max} and calcium refers to the carotid calcium score.

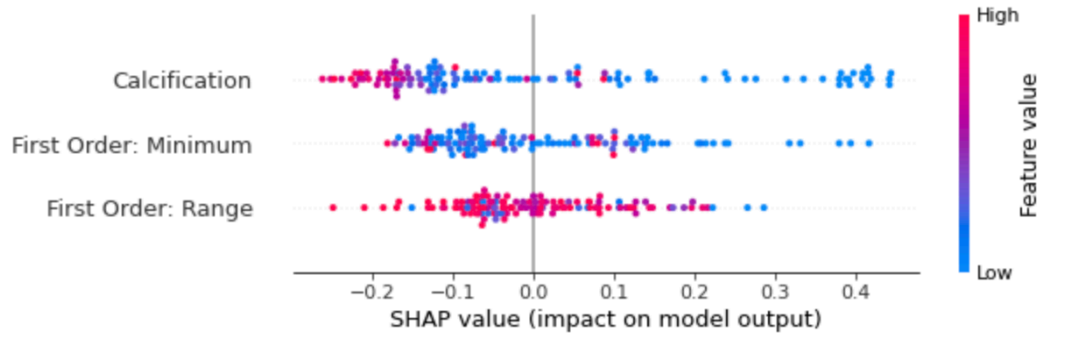
When using the poorly robust radiomic features in the multi-slice (resegmentation) setting, the performance of the radiomic features only were equivalent to chance alone. The random forest classifier with the poorly robust radiomic features and carotid calcium as predictors had the highest performance according to bootstrapping amongst the settings investigated in Table 5.90. SHAP analysis demonstrates that carotid calcification dominates in this model for predicting each of the carotid artery types, see Table 5.91.

Table 5.91 Multi-slice (Resegmentation) Poorly Robust Radiomic Features and Calcium (Random Forest) SHAP analysis for the Training Dataset

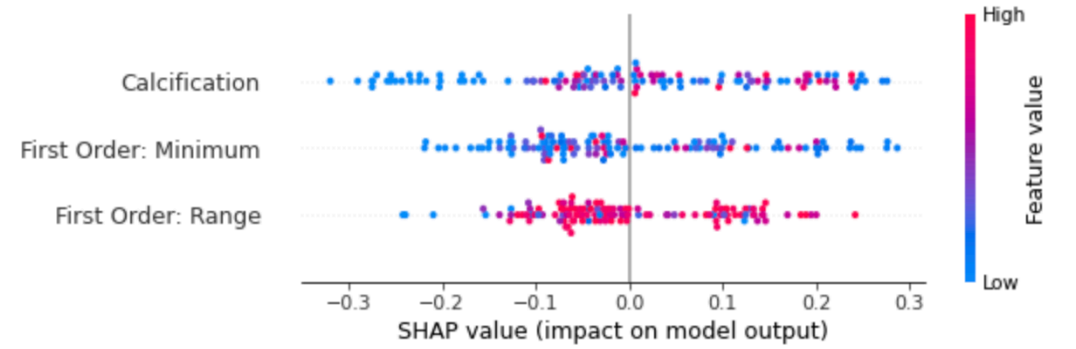
SHAP Feature Importance Plot for the Training Dataset



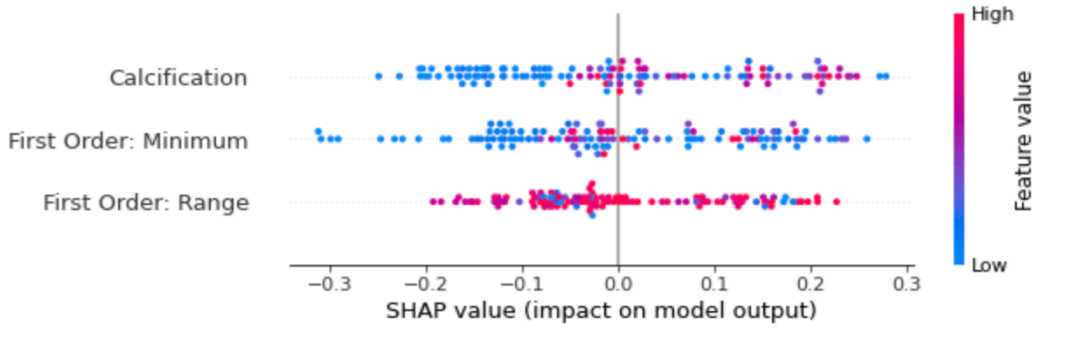
SHAP Summary Value Plot for Asymptomatic carotid arteries



SHAP Summary Value Plot for Culprit carotid arteries



SHAP Summary Value Plot for Non-culprit carotid arteries



Using the external validation dataset, the test accuracy was 50% and the AUC was 0.73. In this setting, the asymptomatic carotid arteries were only identified correctly 50% of the time. However, the culprit carotid artery was correctly identified, but the non-culprit carotid artery was not. The sensitivity and specificity respectively for each class was as follows: Asx (50%, 100%), CC (100%, 92%) and NC (0%, 64%).

Table 5.92 Multi-slice (Resegmentation): External Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**5**	1	4
CC	0	**1**	0
NC	0	1	**0**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the random forest classifier are highlighted in red. Predictors: multi-slice analysis (resegmentation) poorly robust radiomic features and carotid calcification.

5.3.2 Radiomic Feature Correlations with PET

Regarding PET tracer information of the 120 carotid arteries used in the training set of the machine learning classifiers, there was FDG PET information for 120 carotid arteries (40 asymptomatic, 40 culprit and 40 non-culprit carotid arteries), FMISO information for 20 carotid arteries (8 asymptomatic, 6 culprit, 6 non-culprit carotid arteries), NaF information for 44 carotid arteries (22 culprit and 22 non-culprit carotid arteries) and DOTATATE information for 56 carotid arteries (30 asymptomatic, 13 culprit, 13 non-culprit carotid arteries).

5.3.2.1 Multi-slice: Original Image
5.3.2.1.1 FDG mean of TBR_{max}

When considering all 120 carotid arteries together, FDG mean of TBR_{max} was negatively correlated with GLDM: Large Dependence High Grey Level Emphasis (Spearman's r_s = -0.342; p-value = 0.0001) even after correction for multiple comparisons. The other correlations with a p-value<0.05 are listed in Table 5.93.

Table 5.93 Multi-slice (Original Image): FDG Correlations with p-value<0.05

Radiomic Features	Spearman's rho	p-value
Positive Correlations		
GLDM: Low Grey Level Emphasis	0.238	0.009
GLDM: Large Dependence Low Grey Level Emphasis	0.237	0.009
GLRLM: Long Run Low Grey Level Emphasis	0.234	0.010
GLRLM: Low Grey Level Run Emphasis	0.215	0.019
GLRLM: Short Run Low Grey Level Emphasis	0.200	0.029
Negative Correlations		
GLCM: Autocorrelation	-0.204	0.025
GLCM: Joint Average	-0.230	0.012
GLCM: Sum Average	-0.230	0.012
GLSZM: Large Area High Grey Level Emphasis	-0.250	0.006
First Order: Median	-0.250	0.006
GLRLM: Long Run High Grey Level Emphasis	-0.261	0.004
GLSZM: Grey Level NonUniformity	-0.263	0.004

When stratified by carotid artery type there were some negative correlations with a p-value<0.05 for non-culprit carotid arteries. These were: (1) GLRLM: Long Run High Grey Level Emphasis (r_s = -0.397; p-value = 0.011), (2) GLCM: Joint Average (r_s = -0.347; p-value = 0.028), (3) GLCM: Sum Average (r_s = -0.347; p-value = 0.028), (4) GLDM: Large Dependence High Grey Level Emphasis (r_s = -0.345; p-value = 0.029), (5) GLDM: High Grey Level Emphasis (r_s = -0.332; p-value = 0.036), (6) GLCM: Autocorrelation (r_s = -0.328; p-value = 0.039), (7) First Order: Energy (r_s = -0.313; p-value = 0.049) and (8) First Order: Total Energy (r_s = -0.313; p-value = 0.049).

When stratified for culprit carotid arteries, there were no correlations between multi-slice (original image) radiomic features and FDG mean of TBR_{max}.

When stratified for asymptomatic carotid arteries, there were some positive and negative correlations with a p-value<0.05. The positive correlations were: (1) GLCM: Contrast (r_s = 0.329; p-value = 0.038) and (2) GLCM: Difference Average (r_s = 0.319; p-value = 0.045). The negative correlations were: (1) GLCM: Inverse Variance (r_s = -0.373; p-value = 0.018), (2) GLSZM: Zone Variance (r_s = -0.343; p-value = 0.030), (3) GLSZM: Large Area Emphasis (r_s = -0.341; p-value = 0.031), (4) GLRLM: Grey Level NonUniformity (r_s = -0.334; p-value = 0.035), (5) GLSZM: Large Area High Grey Level Emphasis (r_s = -0.328; p-value = 0.039), (6) GLSZM: Grey Level NonUniformity (r_s = -0.317; p-value = 0.046) and (7) GLDM: Grey Level NonUniformity (r_s = -0.316; p-value = 0.047).

5.3.2.1.2 FMISO mean of TBR_{max}

When considering all 20 carotid arteries, there were no statistically significant correlations between multi-slice (original image) radiomic features and FMISO mean of TBR_{max}. After stratifying the correlations by carotid artery type, there were several correlations that remained statistically significant following Benjamini-Hochberg correction for non-culprit carotid arteries. These correlations are shown in Table 5.94.

Table 5.94 Multi-slice (Original): FMISO correlations stratified for non-culprit carotids only

Radiomic Features	Spearman's rho	p-value
Positive Correlations		
First Order: Maximum	1.000	0.000
NGTDM: Strength	1.000	0.000
First Order: Range	0.943	0.005
GLCM: Cluster Prominence	0.943	0.005
GLCM: Cluster Shade	0.943	0.005
GLCM: Cluster Tendency	0.943	0.005
GLCM: Difference Variance	0.943	0.005
GLCM: Idmn	0.943	0.005
GLCM: Idn	0.943	0.005
GLCM: Sum Squares	0.943	0.005
NGTDM: Complexity	0.943	0.005
Negative Correlations		
GLCM: Imc1	-0.943	0.005
NGTDM: Busyness	-0.943	0.005

For culprit carotid arteries, there was a negative correlation with GLSZM: Low Grey Level Zone Emphasis (r_s = -0.886; p-value = 0.019). For asymptomatic carotid arteries, there was a negative correlation with GLSZM: Grey Level NonUniformity (r_s = -0.905; p-value = 0.002) and a positive correlation with GLSZM: Zone Entropy (r_s = 0.738; p-value = 0.037).

5.3.2.1.3 NaF mean of TBR_{max}

When considering all 44 carotid arteries, there were positive and negative correlations between NaF mean of TBR_{max} and multi-slice (original image) radiomic features even after corrections for multiple comparisons. These are listed in Table 5.95. After stratifying by carotid artery status, there was only one correlation that remained statistically significant following Benjamini-Hochberg correction, this was GLDM: Large Dependence Low Grey Level Emphasis (r_s = -0.573; p-value = 0.005) for culprit carotid arteries. There were several correlations for non-culprit and culprit carotid arteries respectively with a p-value<0.05, these are listed in Table 5.96.

Table 5.95 Multi-slice (Original): NaF Correlations - all carotid arteries

Radiomic Features	Spearman's rho	p-value
Positive Correlations		
GLDM: Dependence NonUniformity	0.488	0.0008
GLDM: Dependence Entropy	0.416	0.005
GLRLM: Short Run High Grey Level Emphasis	0.405	0.006
First Order: Entropy	0.403	0.007
GLCM: Joint Entropy	0.400	0.007
GLRLM: Run Length NonUniformity	0.398	0.007
GLRLM: High Grey Level Run Emphasis	0.397	0.008
Negative Correlations		
GLRLM: Grey Level NonUniformity Normalised	-0.402	0.007
GLRLM: Long Run Emphasis	-0.409	0.006
GLRLM: Long Run Low Grey Level Emphasis	-0.410	0.006
First Order: Uniformity	-0.416	0.005
GLDM: Large Dependence Emphasis	-0.419	0.005
GLCM: Joint Energy	-0.424	0.004
GLDM: Large Dependence Low Grey Level Emphasis	-0.431	0.004
GLRLM: Run Variance	-0.435	0.003
GLCM: Maximum Probability	-0.441	0.003
GLDM: Dependence Variance	-0.453	0.002

Table 5.96 Multi-slice (Original): NaF Correlations stratified by carotid artery type

Radiomic Features	Spearman's rho	p-value
Non-culprit Carotid Arteries		
Positive Correlations		
GLDM: Dependence NonUniformity	0.592	0.004
GLSZM: Zone Entropy	0.508	0.016
GLDM: Dependence Entropy	0.495	0.019
GLRLM: Run Length NonUniformity	0.487	0.021
GLSZM: High Grey Level Zone Emphasis	0.465	0.029
First Order: Entropy	0.464	0.030
GLSZM: Size Zone NonUniformity	0.455	0.033
GLCM: Cluster Shade	0.453	0.034
GLSZM: Grey Level Variance	0.443	0.039
GLRLM: Run Percentage	0.441	0.040
First Order: Energy	0.438	0.042
First Order: Total Energy	0.438	0.042
GLCM: Difference Variance	0.438	0.042
GLRLM: Short Run High Grey Level Emphasis	0.425	0.049
Negative Correlations		
GLSZM: Grey Level NonUniformity Normalised	-0.424	0.049
First Order: Minimum	-0.429	0.047
First Order: Uniformity	-0.430	0.046
GLCM: Joint Energy	-0.451	0.035
GLDM: Large Dependence Emphasis	-0.478	0.024
GLRLM: Long Run Emphasis	-0.481	0.024
GLCM: Maximum Probability	-0.503	0.017
GLRLM: Run Variance	-0.511	0.015
GLDM: Dependence Variance	-0.555	0.007
GLDM: Large Dependence High Grey Level Emphasis	-0.580	0.005
Culprit Carotid Arteries		
Positive Correlations		
First Order: Interquartile Range	0.452	0.035
GLRLM: Long Run High Grey Level Emphasis	0.427	0.047
Negative Correlations		
GLCM: Joint Energy	-0.430	0.046
First Order: Uniformity	-0.444	0.038
GLRLM: Grey Level NonUniformity Normalised	-0.465	0.031
GLDM: Low Grey Level Emphasis	-0.465	0.029
GLRLM: Short Run Low Grey Level Emphasis	-0.466	0.029
GLRLM: Low Grey Level Run Emphasis	-0.479	0.024
GLRLM: Long Run Low Grey Level Emphasis	-0.557	0.007
GLDM: Large Dependence Low Grey Level Emphasis	-0.573	0.005

5.3.2.1.4 DOTATATE mean of TBR_{max}

When considering all 56 carotid arteries, there were some positive and negative correlations that remained statistically significant even after Benjamini-Hochberg correct. These are shown in Table 5.97.

Table 5.97 Multi-slice (Original): DOTATATE correlations with all carotid arteries

Radiomic Features	Spearman's rho	p-value
Positive Correlations		
First Order: Skewness	0.440	0.0007
GLCM: Idmn	0.383	0.004
GLCM: Idn	0.377	0.004
Negative Correlations		
GLDM: Large Dependence Emphasis	-0.378	0.004
GLRLM: Run Variance	-0.383	0.003
NGTDM: Contrast	-0.393	0.003
First Order: Median	-0.397	0.002
GLCM: Maximum Probability	-0.401	0.002
GLRLM: Long Run High Grey Level Emphasis	-0.419	0.001
GLDM: Dependence Variance	-0.431	0.0009
GLDM: Large Dependence High Grey Level Emphasis	-0.457	0.0004

After stratifying by carotid artery status, there were correlations that remained statistically significant even after multiple comparisons correction for non-culprit carotid arteries (shown in Figure 5.78), there were no correlations for culprit carotid arteries and there were correlations with a p-value<0.05 for asymptomatic carotid arteries, please see Table 5.98.

Table 5.98 Multi-slice (Original): DOTATATE correlations stratified for asymptomatic carotid arteries only

Radiomic Features	Spearman's rho	p-value
Positive Correlations		
GLRLM: Short Run Emphasis	0.463	0.010
GLRLM: Run Percentage	0.457	0.011
GLRLM: Run Length NonUniformity Normalised	0.446	0.014
GLCM: Joint Entropy	0.364	0.048
Negative Correlations		
GLCM: Idm	-0.362	0.049
GLCM: Id	-0.370	0.044
GLDM: Dependence Variance	-0.370	0.044
GLDM: Large Dependence Emphasis	-0.424	0.019
GLRLM: Run Variance	-0.451	0.012
GLRLM: Long Run Emphasis	-0.459	0.011

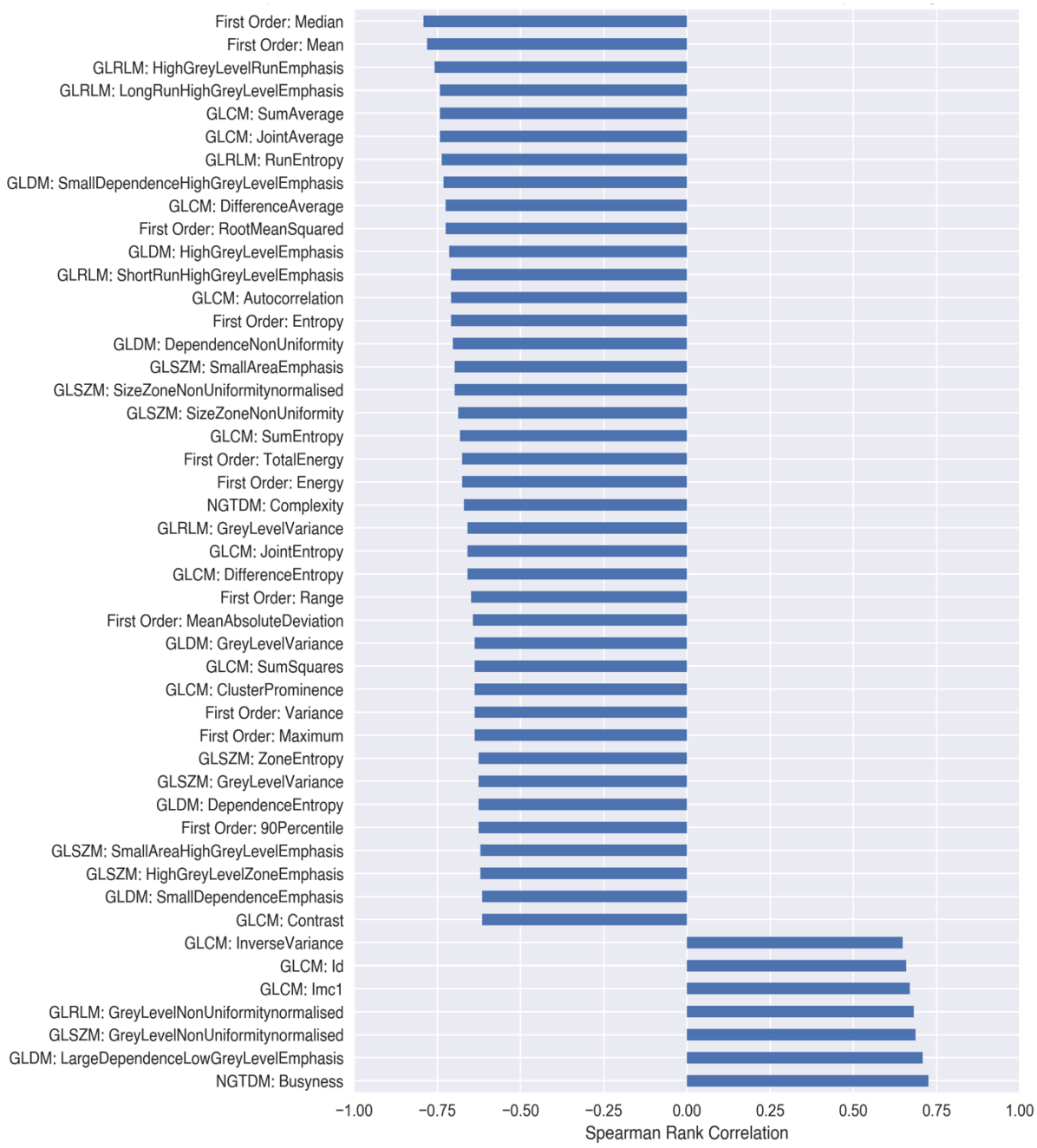


Figure 5.78 Spearman Rank Correlation of DOTATATE with Multi-slice (Original) Radiomic Features - Non-culprit carotid arteries only

5.3.2.2 Multi-slice: Resegmentation

5.3.2.2.1 FDG mean of TBR_{max}

When considering all 120 carotid arteries, there were weak positive and weak negative correlations between FDG mean of TBR_{max} and second order radiomic features. The correlations that remained statistically significant following correction for multiple comparisons are demonstrated in Figure 5.79.

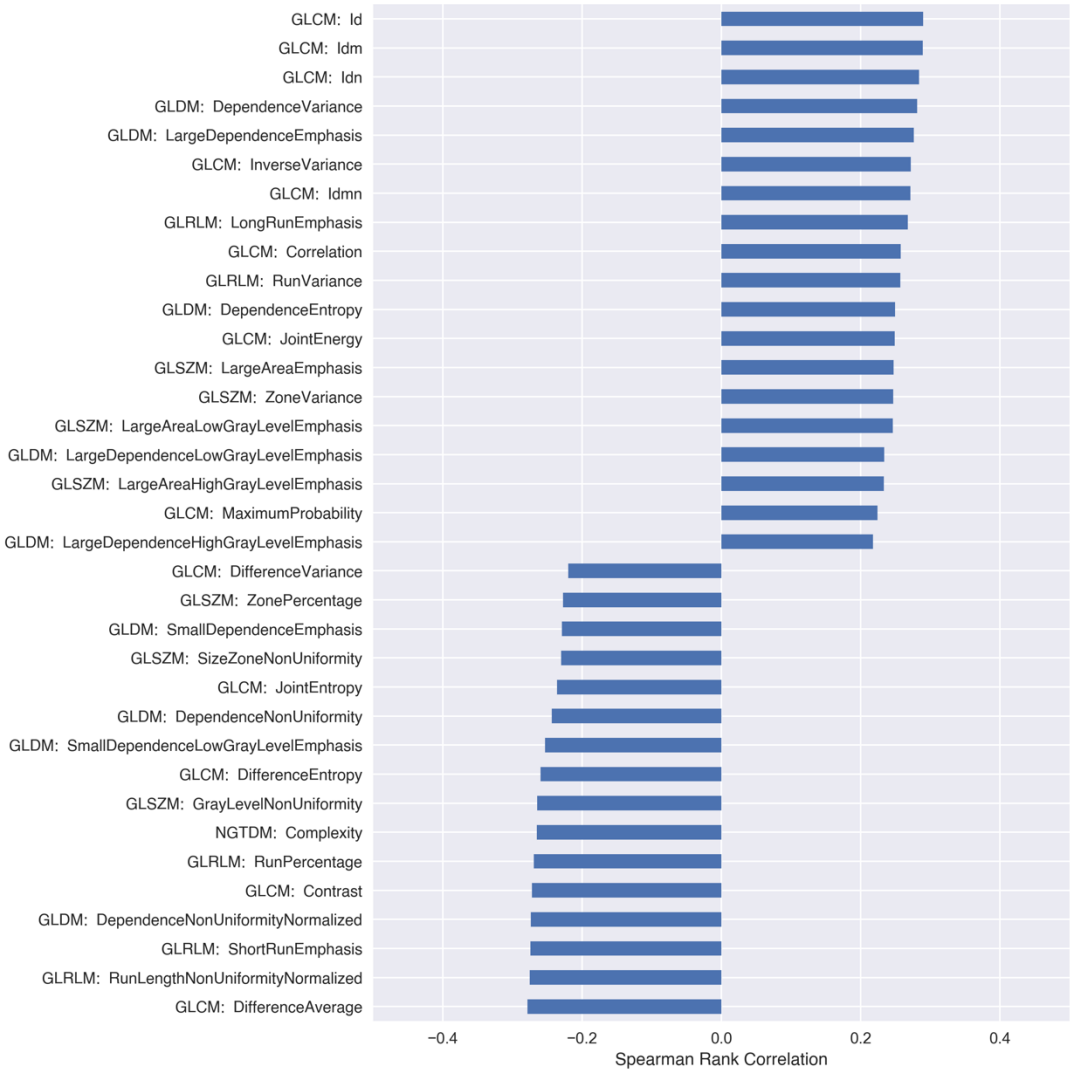


Figure 5.79 Spearman Rank Correlation of FDG mean of TBR_{max} with Radiomic Features (Resegmentation)

These correlations were no longer there when stratified by carotid artery type for asymptomatic carotid arteries only, and culprit carotid arteries only respectively. For the non-culprit carotid arteries, only two radiomic features were weakly negatively correlated with FDG mean of TBR_{max}: (1) GLSZM: Size Zone NonUniformity (r_s = -0.343; p-value = 0.030) and (2) GLSZM: Size Zone NonUniformity Normalised (r_s = -0.326; p-value = 0.040).

5.3.2.2.2 FMISO mean of TBR_{max}

Following resegmentation, when all 20 carotid arteries were considered, there were moderate positive and negative correlations between FMISO mean of TBR_{max} and second order radiomic features. Those that remained statistically significant following Benjamini-Hochberg correction are shown in Figure 5.80.

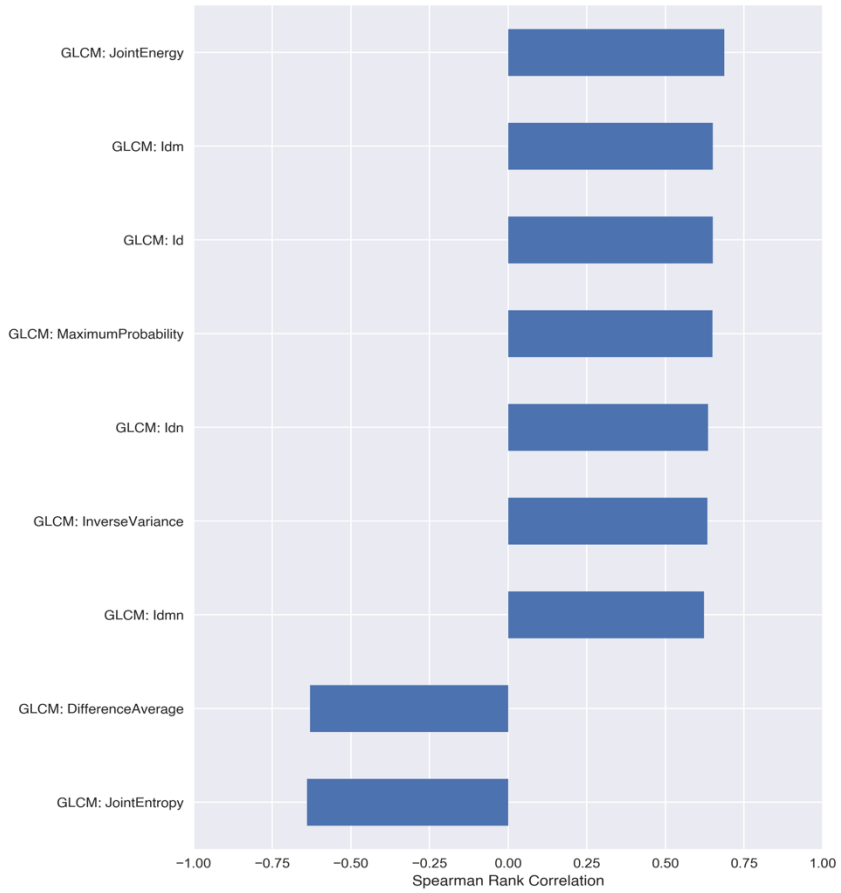


Figure 5.80 Spearman Rank Correlation of FMISO mean of TBR_{max} with Radiomic Features (Resegmentation)

When stratified according to carotid artery type, there were no correlations for non-culprit carotid arteries only. For culprit carotid arteries, there were some negative correlations between FMISO TBR_{max} with a p-value<0.05: (1) GLDM: Dependence NonUniformity (r_s = -0.886; p-value = 0.019), (2) GLRLM: Run Length NonUniformity (r_s = -0.829; p-value = 0.042), (3) GLSZM: Grey Level Variance (r_s = -0.886; p-value = 0.019) and (4) GLSZM: Zone Entropy (r_s = -0.886; p-value = 0.019). There was also a strong positive correlation with NGTDM: Strength (r_s = 0.829; p-value = 0.042).

For asymptomatic carotid arteries, there were some strong correlations that remained statistically significant after multiple comparisons correction, these were: (1) GLDM: Large Dependence Low Grey Level Emphasis (r_s = 0.952; p-value = 0.0003), (2) GLSZM: Large Area Low Grey Level Emphasis (r_s = 0.929; p-value = 0.0009), (3) GLCM: Id (r_s = 0.905; p-value = 0.002), (4) GLCM: Idm (r_s = 0.905; p-value = 0.002), (5) GLDM: Dependence Variance (r_s = 0.905; p-value = 0.002) and (6) GLDM: Dependence NonUniformity Normalised (r_s = -0.905; p-value = 0.002).

5.3.2.2.3 NaF mean of TBR_{max}

Following resegmentation, when considering all 44 carotid arteries, NaF mean of TBR_{max} was weakly positively correlated with GLSZM: Size Zone NonUniformity (r_s = 0.414; p-value = 0.005) and negatively correlated with (1) GLCM: Joint Energy (r_s = -0.372; p-value = 0.013) and (2) GLRLM: Grey Level NonUniformity Normalised (r_s = -0.367; p-value = 0.014).

When stratified according to carotid artery type, there were some correlations with a p-value <0.05 for non-culprit carotid arteries. These were: (1) GLCM: Joint Entropy (r_s = 0.474; p-value = 0.026), (2) GLSZM: Size Zone NonUniformity (r_s = 0.449; p-value = 0.036), (3) First Order: Entropy (r_s = 0.441; p-value = 0.040), (4) GLCM: Joint Energy (r_s = -0.464; p-value = 0.030), (5) GLCM: Maximum Probability (r_s = -0.427; p-value = 0.047) and (6) First Order: Uniformity (r_s = -0.426; p-value = 0.048).

For culprit carotid arteries, there was GLDM: Small Dependence Low Grey Level Emphasis (r_s = 0.439; p-value = 0.041). No asymptomatic carotid arteries had NaF PET scans.

5.3.2.2.4 DOTATATE mean of TBR_{max}

Following resegmentation, when considering all 56 carotid arteries, there were moderate positive and negative correlations between DOTATATE mean of TBR_{max} and first and second order radiomic features. Those that remained statistically significant following Benjamini-Hochberg correction are shown in Figure 5.81.

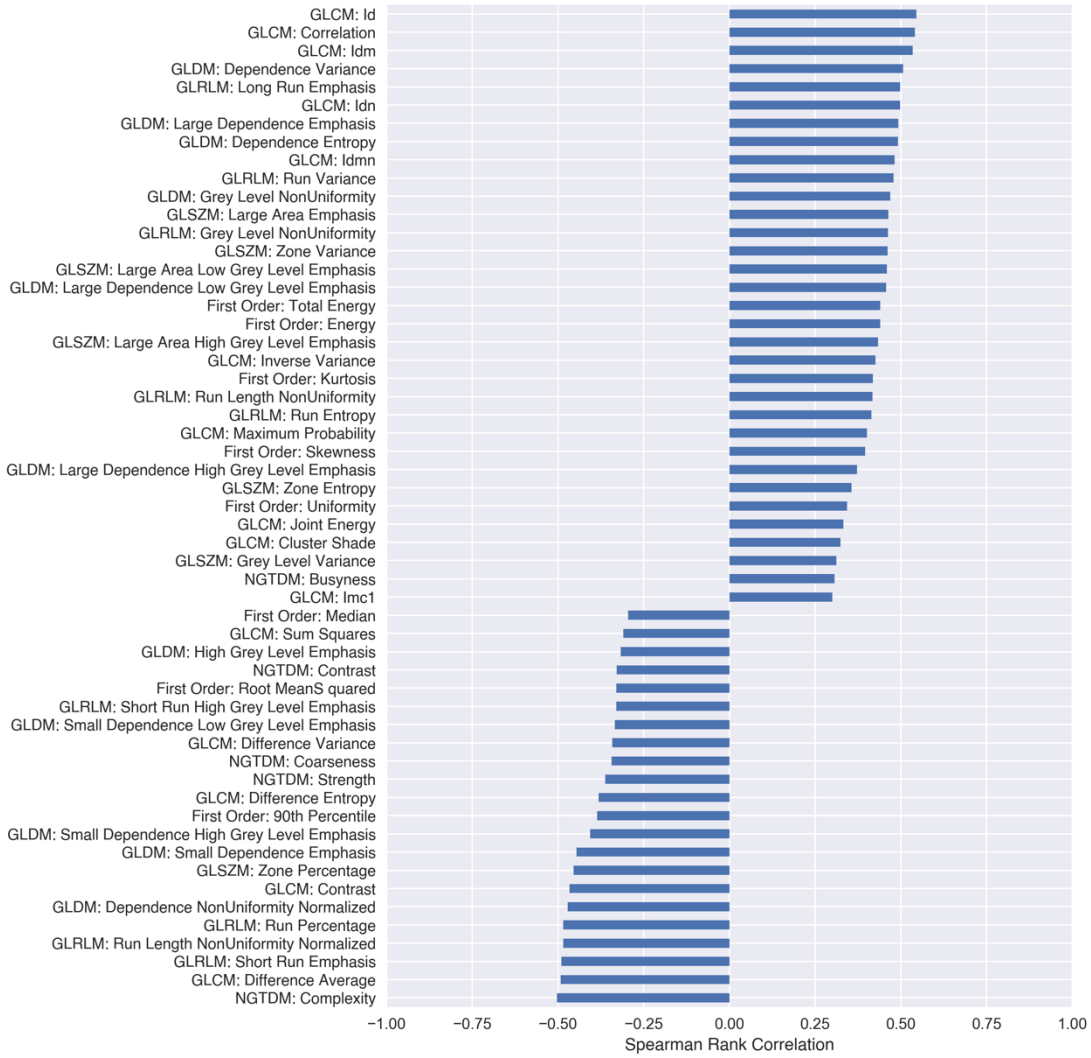


Figure 5.81 Spearman Rank Correlation of DOTATATE mean of TBR_{max} with Radiomic Features (Resegmentation)

When stratified according to carotid artery type, there were 5 strong correlations following multiple comparisons correct for non-culprit carotid arteries. These were: (1) GLCM: Maximum Probability (r_s = 0.881; p-value = 0.0002), (2) GLCM: Joint Energy (r_s = 0.867; p-value = 0.0003), (3) GLDM: Dependence Variance (r_s = 0.839; p-value = 0.0006), (4) GLRLM: Run Length NonUniformity Normalised (r_s = -0.769; p-value = 0.003) and (5) GLRLM: Short Run Emphasis (r_s = -0.769; p-value = 0.003).

For culprit carotid arteries, there were no correlations that remained statistically significant following multiple comparison corrections. The positive correlations with a p-value<0.05 were: (1) NGTDM: Complexity (r_s = 0.615; p-value = 0.033), (2) GLDM: Dependence NonUniformity Normalised (r_s = 0.643; p-value = 0.024), (3) GLCM: Contrast (r_s = 0.678; p-value = 0.015), (4) GLCM: Difference Average (r_s = 0.685; p-value = 0.014), (5) GLCM: Difference Entropy (r_s = 0.685; p-value = 0.014), (6) GLCM: Dependence NonUniformity (r_s = 0.699; p-value = 0.011) and (7) GLCM: Difference Variance (r_s = 0.727; p-value = 0.007). The negative correlations with a p-value<0.05 were: (1) GLCM: Inverse Variance (r_s = -0.699; p-value = 0.011), (2) GLCM: Idmn (r_s = -0.685; p-value = 0.014), (3) GLCM: Idn (r_s = -0.685; p-value = 0.014), (4) GLCM: Id (r_s = -0.678; p-value = 0.015), (5) GLCM: Idm (r_s = -0.678; p-value = 0.015), (6) GLDM: Large Dependence High Grey Level Emphasis (r_s = -0.664; p-value = 0.018) and (7) GLCM: MCC (r_s = -0.594; p-value = 0.042).

There were no correlations when stratified for asymptomatic carotid arteries only.

5.3.3 Radiomic Feature Correlations with Immunohistochemistry

7 symptomatic patients had histology with immunohistochemical staining of culprit carotid plaque for the presence of macrophages and the percentage area of CD68+ was quantified.

5.3.3.1 Single-slice: Original Image

Of the 93 radiomic features extracted from CTA (original image), %CD68+ was negatively correlated with First Order: Minimum (Spearman's r_s = -0.893; p-value = 0.007) and negatively correlated with GLDM: Small Dependence Low Grey Level Emphasis (r_s = -0.893; p-value = 0.007).

The correlations of all 93 radiomic features with %CD68+ and associated p-values are summarised in Figure 5.82.

5.3.3.2 Single-slice: Resegmentation

Of the 93 radiomic features extracted from CTA (with resegmentation), %CD68+ was negatively correlated with First Order: Minimum (r_s = -0.883; p-value = 0.008), positively correlated with First Order: Range (r_s = 0.893; p-value = 0.007), positively correlated with GLDM: Dependence Entropy (r_s = 0.857; p-value = 0.014), positively correlated with GLRLM: Grey Level Variance (r_s = 0.821; p-value = 0.023) and positively correlated with GLSZM: Grey Level Variance (r_s = 0.964; p-value = 0.0005).

The correlations of all 93 radiomic features (extracted following resegmentation) with %CD68+ and their associated p-values are summarised in Figure 5.83.

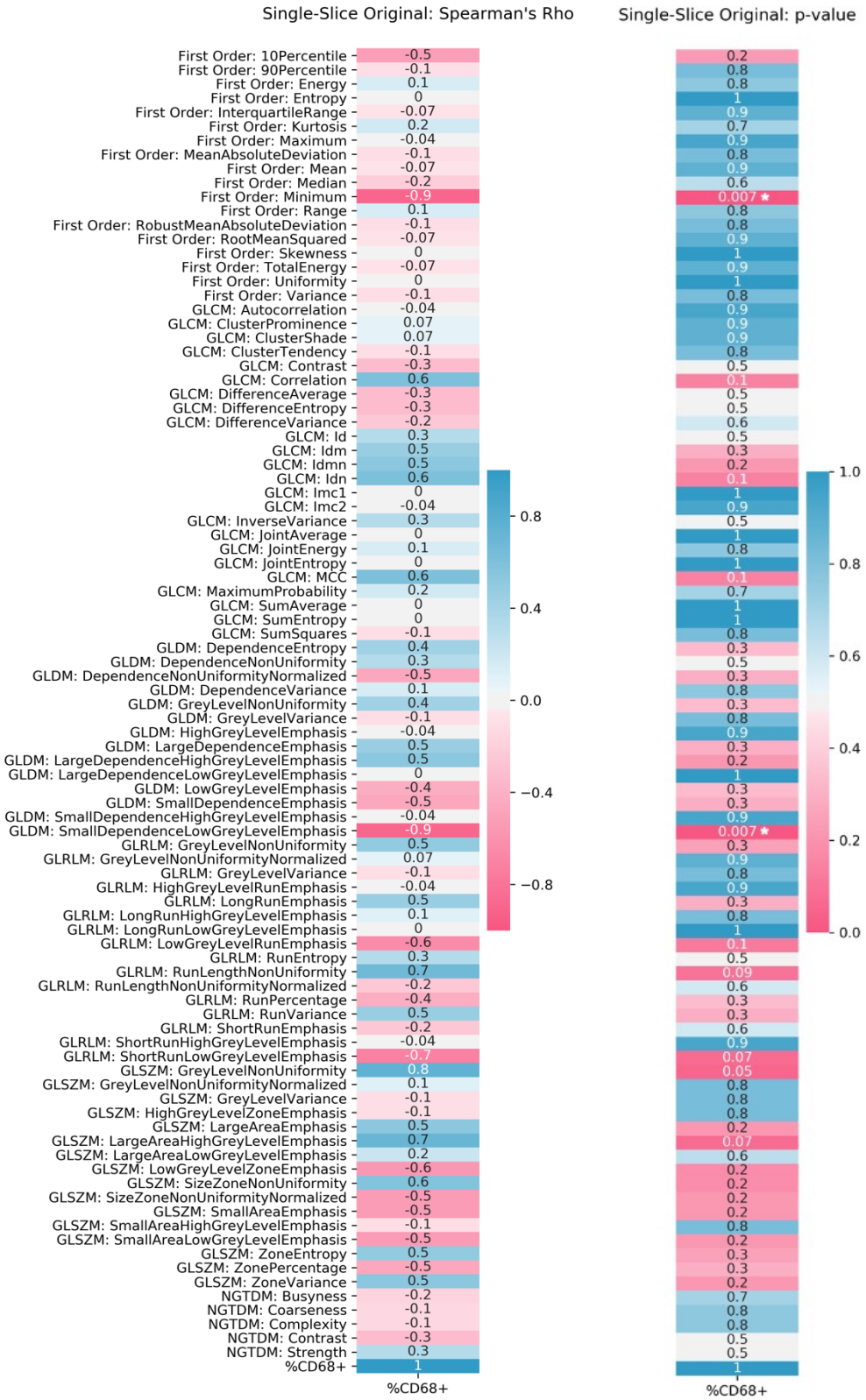


Figure 5.82 CTA Radiomic Feature (Original Image) Correlations with %CD68+

*Spearman Rank Correlations were calculated between 93 radiomic features and the %CD68+ of culprit carotid plaque immunohistochemically stained sections. The numbers in each cell represent Spearman's rho on the left and the associated p-value on the right; * $p<0.05$*

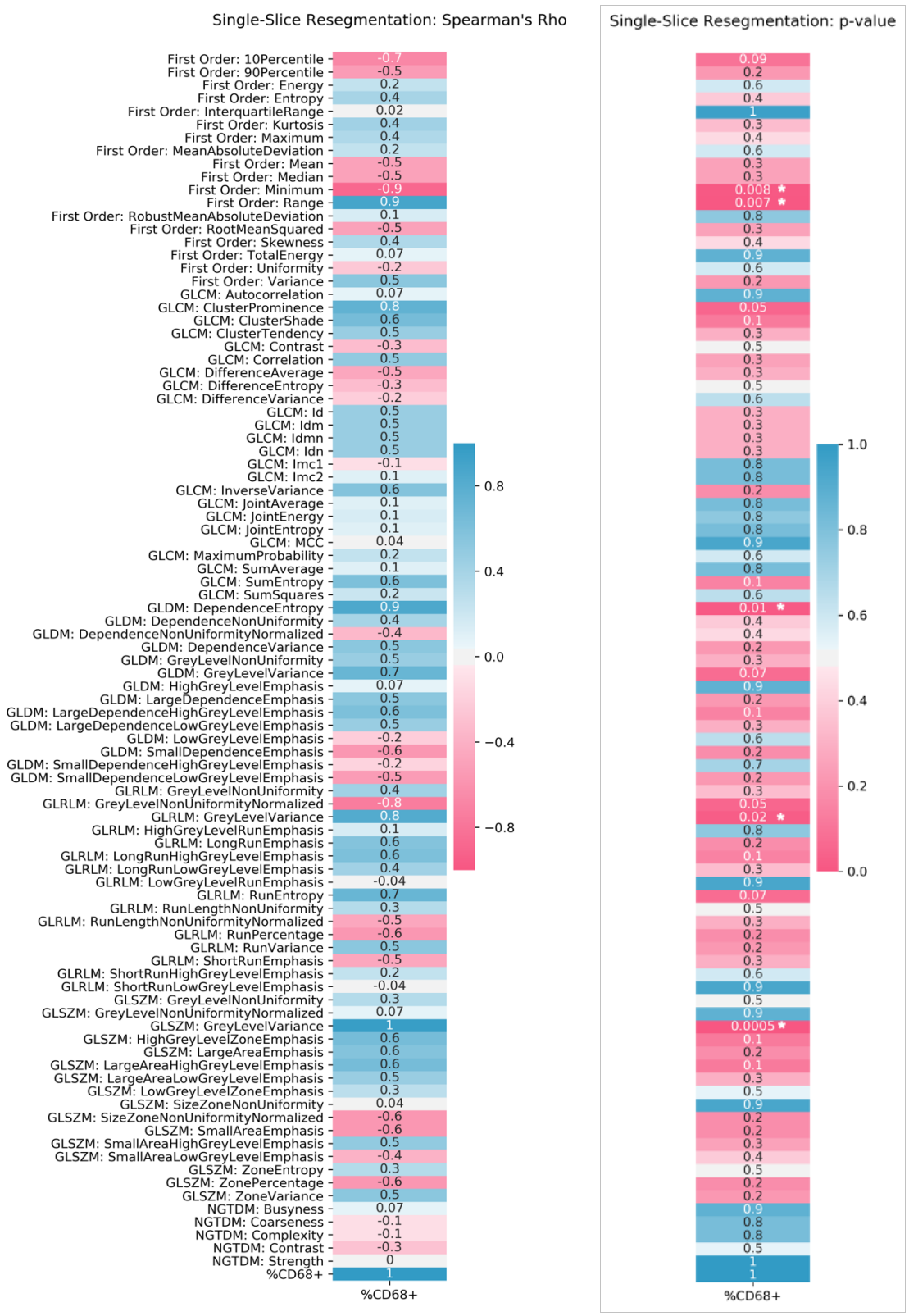


Figure 5.83 CTA Radiomic Feature (Resegmentation) Correlations with %CD68+

*Spearman Rank Correlations were calculated between 93 radiomic features and the %CD68+ of culprit carotid plaque immunohistochemically stained sections. The numbers in each cell represent Spearman's rho on the left and the associated p-value on the right; * $p<0.05$*

5.4 Discussion

5.4.1 Summary of key findings

In this chapter, we investigated the ability of the highly robust radiomic features (identified in Chapter 4) to classify carotid arteries into 3 classes: culprit, non-culprit or asymptomatic. We compared the predictive performance of these highly robust features with the performance of 1) using all radiomic features as predictors and 2) using poorly robust features only to assess the assumption that robust features improve model generalisability. In addition, we tested the impact on predictive ability of integrating radiomic features with PET data and carotid calcium scores. SHAP analysis was used to interpret the machine learning models. Finally, we aimed to investigate the possible biological basis of carotid artery radiomic features by assessing their correlations with PET data and immunohistochemistry.

We found that highly robust radiomic features (whether extracted from the original image or from the re-segmented image: [0, 200] HU) outperformed FDG PET mean of TBR_{max} and carotid calcium score as predictors for multi-class classification. There was little to no improvement with the addition of PET information to radiomic features. When carotid calcification was combined with radiomic features (original image), there was no improvement in predictive performance. However, as seen in Chapter 4, combining carotid calcium with radiomic features (following resegmentation) provided complementary information that did improve carotid artery differentiation. Whilst using all radiomic features as predictors led to similar performances as using the highly robust radiomic features, using the poorly robust radiomic features led to performances worse than carotid calcification alone.

The biological correlation analyses in this chapter revealed associations between radiomic features and different PET tracers: FDG, FMISO, NaF and DOTATATE that measure various aspects of inflammation. There were also tentative associations with %CD68+ (a marker of macrophage presence) as determined by immunohistochemistry.

5.4.2 Interpretation of findings

5.4.2.1 Multi-class Feature-based Machine Learning
5.4.2.1.1 Multi-slice (Original) Radiomic Features

SHAP analysis revealed that when using highly robust radiomic features (original image) as predictors, asymptomatic carotid arteries were associated with higher values for GLDM: Large Dependence High Grey Level Emphasis, First Order: Root Mean Squared and GLCM: Joint Average. However, they were associated with lower values of First Order: Maximum and First Order: Robust Mean Absolute Deviation. This may partly reflect how asymptomatic carotid arteries were more likely to have less carotid stenosis and therefore increased contrast filling in the lumen, as well as decreased carotid calcification compared with the symptomatic carotid arteries. This agrees with Chapter 3 findings in which certain radiomic features extracted without resegmentation (i.e. original image) were related to carotid calcification and carotid stenosis.

Culprit carotid arteries were associated with lower values of First Order: Root Mean Squared, GLDM: Large Dependence High Grey Level Emphasis and GLCM: Joint Average but higher values of First Order: Robust Mean Absolute Deviation and First Order: 90^{th} Percentile, potentially reflecting the propensity of culprit carotid arteries to have higher degrees of carotid stenosis and carotid calcification. Interestingly, non-culprit carotids were associated with lower values of First Order: 90^{th} percentile and GLDM: Large Dependence High Grey Level Emphasis even though the carotid calcium burden between culprit versus non-culprit carotid arteries are equivalent, suggesting that the radiomic features may be capturing other aspects of the carotid artery besides calcium load and luminal stenosis. For example, this could include characteristics of the underlying carotid plaque.

When using radionic features extracted from the multi-slice (original image) setting, carotid calcification provided little to no benefit to the predictive ability of the radiomic features. This is because the radiomic features associated with high grey level values such as First Order: Maximum already captured carotid calcification information. Using the highly robust radiomic features only versus using all radiomic features produced similar predictive performances according to cross-validation or bootstrapping. However, the performance of using highly robust radiomic features only was higher than using all radiomic features on the external validation set (AUC of 0.86 versus 0.73). This may reflect the curse of dimensionality which

supports the use of fewer explanatory variables for a simpler model with better estimations[218], especially in cases where the number of predictors is larger than the number of samples. In addition, this finding supports the underlying assumptions of Chapter 4's robustness analyses i.e. the incorporation of only robust radiomic features into predictive models improves their ability to perform on new image datasets, referred to as 'good generalisability'.

5.4.2.1.2 Multi-slice (Resegmentation) Radiomic Features

In multi-class classification, as opposed to binary classification of culprit versus non-culprit carotid arteries (Chapter 4), resegmentation did not additionally benefit the predictive performance of the machine learning classifiers when only using radiomic features as predictors. Using the radiomic features extracted in the multi-slice (resegmentation) setting produced similar results to using multi-slice (original image) radiomic features as predictors. This may reflect how it is easier to differentiate carotid arteries from asymptomatic patients and symptomatic patients compared with differentiating culprit versus non-culprit carotid arteries as there are greater differences in the presence of carotid atherosclerotic plaque, carotid calcium burden and carotid artery stenosis for example.

When using only the highly robust radiomic features, asymptomatic carotid arteries were associated with lower values of GLDM: Dependence Variance and GLDM: Large Dependence High Grey Level Emphasis. They were associated with higher levels of GLDM: Difference Variance and GLRLM: Long Run High Grey Level Emphasis. Conversely, culprit carotid arteries were associated with lower GLRLM: Long Run High Grey Level Emphasis and higher GLDM: Dependence Variance which could potentially reflect low attenuation values associated with increased carotid stenosis, the presence of non-calcified plaque or a lipid-rich core. Non-culprit carotid arteries had a radiomic profile in between asymptomatic and culprit carotid arteries with moderate to high values of GLDM: Dependence Variance and GLCM: Difference Variance.

In the context of resegmentation, carotid calcification improved predictive ability. This firstly reflects how radiomic features extracted following resegmentation to [0, 200] do not contain much information about carotid calcium. Secondly, it demonstrates that carotid calcium and the information captured by radiomic features (with resegmentation) contain complementary information that can be exploited with multivariable models.

5.4.2.1.3 Poorly Robust Radiomic Features

The poorly robust radiomic features previously identified in Chapter 4 were related to low attenuation values such as GLSZM: Low Grey Level Zone Emphasis. These values likely related to carotid artery perivascular fat which would have been the main contributor to inter-observer variation in carotid artery segmentations as the boundary between the outer adventitia and the extra-vessel matrix was a subjective decision. Surprisingly, using the 10 poorly robust radiomic features extracted in the multi-slice (original image) setting, there was good predictive performance. The performance on external validation was AUC of 0.88 which suggests there was useful radiomic information related to the low attenuation values of the carotid artery (which related to perivascular fat). This is further supported in how the poorly robust features extracted in multi-slice analysis with resegmentation did not perform well. Resegmentation provides a lower limit of 0 HU, which would exclude information about perivascular fat from being captured by the radiomic features. In coronary arteries, phenotypic changes in perivascular fat are associated with atherosclerotic inflammation as a result of pro-inflammatory cytokine release[219]. Radiomic analysis of pericoronary adipose tissue has recently been shown to reflect coronary inflammation[97] with the ability to identify patients with acute myocardial infarction from asymptomatic patients[220]. Therefore, a future direction for carotid artery radiomics studies could look into the value of incorporating pericarotid adipose tissue into the region-of-interest.

5.4.2.1.4 Radiomic Features versus PET in Multi-class Classification

Whether using only highly robust radiomic features, all radiomic features or only poorly robust radiomic features as predictors for the multi-class classification of asymptomatic, culprit or non-culprit carotid arteries, carotid CTA radiomic features always outperformed FDG PET TBR_{max} alone (which performed similarly to chance). Although FDG is associated with imaging carotid atherosclerotic inflammation, there have been conflicting results[203] regarding its ability to discriminate between culprit and non-culprit atheroma. This may reflect the non-specific nature of FDG uptake, particularly when compared to the higher specificity of NaF^{151} uptake or $DOTATATE^{152}$ for carotid atherosclerosis. As such, it was not entirely surprising that FDG PET was not a great predictor in multi-class classification.

When PET information was incorporated along with the radiomic features as predictors, no additional benefit was provided to the radiomic features. This suggests that carotid CTA radiomics may have the potential to replace FDG PET for some applications in atherosclerotic imaging as it is better able to discriminate between different carotid artery types. Not all of the carotid arteries in this study had corresponding PET data. Different PET tracer information was available to different extents for the carotid CTA dataset, with FDG PET being the most abundant. This is why FDG PET was used as the main PET comparator in multi-class classification in this chapter. Future work should compare the predictive ability of radiomic features with other PET tracers that are more specific to carotid atherosclerosis (such as NaF and DOTATATE), as well as investigate the value of an integrated cross-modality approach. In addition, future work could investigate changes in radiomic markers with therapy.

5.4.2.1.5 Machine Learning Classifiers

In this chapter, we investigated several machine learning classifiers. This approach acknowledged the "no free lunch" theorem[221], which describes how there is no universal best model for every task. This no 'one size fits all' approach is well recognised in the radiomics field and testing cross-combinations of ML algorithms is common practice[222,223]. Overall, the machine learning classifiers evaluated achieved similar predictive performances (random forest, LASSO, Elastic Net, neural network and XGBoost), with the exception of the decision tree classifier which consistently had the worst performance. The decision tree classifier is the simplest amongst those investigated and has a higher risk of overfitting on the training data than the other classifiers due to its simplistic nature[189]. Random forests are an extension of the decision tree classifier as it creates an ensemble from different decision trees generated on randomly selected data samples. This helps to mitigate the overfitting problem as the average of all the predictions from individual decision trees should cancel out the biases and improve classification accuracy[189,194]. Overall, depending on the predictors used, the random forest classifier, LASSO and Elastic Net logistic regression-based classifiers or the neural network classifier achieved the highest performance. This highlights the importance of assessing several ML classifiers in radiomic workflows.

5.4.2.2 Radiomic Feature Correlations with PET
5.4.2.2.1 Correlations with FDG

FDG is a radiolabelled glucose analogue that is taken up by all cells that metabolise glucose, including active macrophages in carotid atherosclerotic plaque[203]. A recent meta-analysis found that the FDG signal correlated with carotid atherosclerotic inflammation[203]. In general, there was a positive correlation between FDG and radiomic features associated with low attenuation values such as GLDM: Low Grey Level Emphasis. The GLDM: Low Grey Level Emphasis measures the distribution of low grey level values whereby higher values reflect a greater concentration of low grey level values in the image. This suggests that with greater carotid inflammation, there is an associated greater concentration of low CT attenuation values. This could reflect increased carotid stenosis, increased adiposity within or around the carotid artery (if captured by the ROI) or the presence of low-attenuation plaque for example, which is a high-risk feature on coronary CT angiography[224]. Conversely, there were weak negative correlations with radiomic features associated with high attenuation values such as GLDM: Large Dependence High Grey Level Emphasis. This suggests that with greater carotid inflammation there was a tendency to more disjunction between carotid artery components with high CT attenuation values. This could reflect the presence of mixed plaque for example as opposed to a calcified plaque with a large mass of macrocalcification.

In oncology, CT texture features have been correlated with non-small cell lung cancer FDG uptake: (1) negative correlation with First Order: Uniformity and (2) positive correlation with First Order: Mean Grey-level Intensity[225]. However, it may be difficult to directly compare CT texture features in one study to another due to differences in radiomic feature definitions, scanning parameters and disease phenotype. Nevertheless, the results of this chapter suggest that carotid CTA radiomics have a plausible basis and merit further investigation.

5.4.2.2.2 Correlations with FMISO

FMISO is a common PET radiotracer used for hypoxia imaging[153,226]. Higher FMISO uptake has been observed in culprit versus non-culprit atheroma and correlated with FDG uptake, consistent with hypoxia as either a contributing factor to inflammation or playing a direct role in FDG uptake[153].

In computer graphics, a primitive is defined as a basic image element such as a line, a square or a polygon which combines together to create higher-level, more complex objects[227]. FMISO was strongly positively correlated with NGTDM: Strength where a high value of NGTDM: Strength suggests that the primitives within an ROI are distinct and easily identifiable with 'large coarse differences in grey level intensities'[188]. This suggests that an increased FMISO signal, which is used as a marker of hypoxia, is associated with carotid arteries that have several distinct components. FMISO was also positively correlated with NGTDM: Complexity which indicates when there are many primitive components within an ROI i.e. the ROI is non-uniform. This idea of non-uniformity is reflected in the strong positive correlations with First Order: Range and GLCM: Cluster Prominence, where a higher value suggests greater asymmetry about the mean grey level value. This is further supported by the strong negative correlation with GLCM: Imc1, which is the informational measure of correlation. The lower the value, the less mutual information there is between different pixel values, with a value of 0 indicating no mutual information.

In the literature, First Order: Standard Deviation has been found to positively correlate with hypoxia in lung cancer[76]. This suggests that increased non-uniformity may reflect underlying hypoxia not only in oncology, but also in carotid atherosclerosis.

5.4.2.2.3 Correlations with NaF

NaF was previously found to be a marker of increased microcalcification which is associated with greater plaque vulnerability[151]. NaF information was only available for symptomatic patients and therefore only culprit and non-culrpit carotid arteries. Overall, there were moderate positive correlations between NaF and measures of CT signal heterogeneity such as GLDM: Dependence NonUniformity and First Order: Entropy and negative correlations with radiomic features related to high attenuation values such as GLDM: Large Dependence Low Grey Level

Emphasis as well as First Order: Uniformity. This suggests that the more CT signal heterogeneity is present within the ROI, the greater the level of microcalcification within the carotid artery plaque.

5.4.2.2.4 Correlations with DOTATATE

DOTATATE targets somatostatin receptor subtype-2 that are present in proinflammatory macrophages, allowing for greater cell specificity than FDG in imaging atherosclerotic inflammation[152]. Interestingly DOTATATE was positively correlated with GLCM: Idmn and GLCM: Idn (inverse difference moment normalised and inverse difference normalised, respectively) which are both measures of local homogeneity within an ROI. Local homogeneity could be present in carotid plaques that are completely calcified for example, or completely noncalcified. DOTATATE was negatively correlated with radiomic features associated with high attenuation values such as GLDM: Large Dependence High Grey Level Emphasis and GLRLM: Long Run High Grey Level Emphasis suggesting that the greater the discontinuity in high grey level values, the greater the level of carotid inflammation.

5.4.2.2.5 Overall interpretation of PET correlations with Radiomic Features

There is a growing evidence base that radiomic features are associated with biologically relevant disease processes. This is greatest in the field of oncology, whereby CT radiomic features have been associated with histopathologic correlates such as tumour grade or pathophysiological processes such as hypoxia[68]. More recently, CT radiomic features have been associated with circulating tumour DNA concentrations, which may be used for routine treatment response monitoring[228]. In this chapter, our preliminary study suggests that there is also a biological basis to carotid CTA radiomic features. This is related to carotid inflammation, a central unifying concept to the pathophysiology of atherosclerosis and malignancy.

5.4.2.3 Radiomic Feature Correlations with Immunohistochemistry

The radiomic features were correlated with quantitative immunohistochemical staining for macrophage density (CD68) of the corresponding endarterectomy specimens. One of the histological hallmarks of the ‘vulnerable plaque’ (discussed in Chapter 1, section 1.1.2) is characterised by extensive macrophage infiltration. Although the sample size for carotid histology specimens was limited in this study, there were some strong correlations between

%CD68+ (a marker of macrophage infiltration) and carotid CTA radiomic features. %CD68+ was negative correlated with radiomic features related to low CT attenuation values which could reflect increased carotid stenosis, the presence of non-calcified plaques, lipid-rich atherosclerotic plaque or aspects of carotid perivascular fat (if that was captured within the ROI).

On the other hand, %CD68 was positively correlated with different measures of CT signal heterogeneity such as GLDM: Dependence Entropy or First Order: Range following resegmentation to [0, 200] HU. This suggests that increased macrophage infiltration in culprit carotid arteries is associated with increased dispersion of the grey levels between 0-200 HU. These findings suggest that certain radiomic features correlate with indicators of plaque vulnerability. This supports the underlying assumption of the radiomics field i.e. "biomedical images contain information of disease-specific processes"[73].

5.4.3 Limitations and Future Work

In this study, the default configurations (in the Python sklearn and xgboost packages) for the machine learning classifiers were used. Although using the default values for the machine learning classifiers was a valid approach[229], hyperparameter tuning could have further improved the predictive performance achieved. This would have required either manual configuration of the hyperparameter decisions or the use of a hyperparameter tuning strategy[230] and could be explored in future work.

For multi-class classification, we investigated three approaches: (1) using highly robust radiomic features only, (2) using all radiomic features and (3) using poorly robust radiomic features only. For the approaches where a subset of the original radiomic feature set was used, the curse of dimensionality[231] was mitigated as fewer features were used as predictors for the relatively limited size of the training dataset. However, for the approach that used all 93 radiomic features, optimal performance was likely not achieved as there could have been the issue of high dimensionality and multicollinearity, particularly without an additional feature selection or dimensionality reduction step employed. No feature selection or feature engineering step was used in order to make the three approaches directly comparable with each other.

The external validation set used in this study comprised 12 carotid arteries, 10 of which were asymptomatic, with only 1 culprit and 1 non-culprit carotid artery. This was a limited number of samples and the distribution of carotid artery classes was different to the dataset that the model was trained on (training data had 40 asymptomatic, 40 culprit and 40 non-culprit carotid arteries). However, the distribution of carotid artery types in the external validation set was akin to that in a general screening of the population scenario and despite this discrepancy in class distribution, good predictive performance was achieved using the radiomic features. Another caveat is that the external validation set consisted of the scans from patients that did not receive a corresponding FDG PET scan which could have biased the dataset in some way. Furthermore, the robustness analyses in Chapter 4 identified the radiomic features robust to ROI perturbations that mimicked possible over- and under-estimations of carotid artery segmentations as occurs with *inter-rater* variability. However, the radiomic features used as predictors in this external validation set were generated from ROIs drawn by a single reader (EPVL) and originated from carotid CTA scans captured from the same centre and scanner as the training data. Therefore, we only managed to capture a glimpse into the generalisability of the robust radiomic features. Future work should aim towards acquiring a separate, independent external carotid CTA dataset with ROIs drawn by an independent reader.

For the biological validation of the radiomic features, not all of the carotid arteries had corresponding PET information from all the different PET tracers since the dataset was curated from three separate imaging studies: VISION[152], CHAI[153] and ICARUSS[151] and thus certain correlation analyses were limited by the available sample size of carotid arteries. Regarding the histology samples, we were limited to carotid plaque specimens from only 7 culprit carotid arteries. In addition, the histology samples were of insufficient quality for radiomic analysis of the histology slide or to individual count the macrophages present in the carotid artery plaque. Furthermore, the radiomic features in this study were not limited to the carotid plaque alone but had contributions from all areas of the carotid artery up to the outer carotid adventitia. Future work could investigate radiomic analysis of radiomic analysis confined specifically to the carotid plaque as this would permit direct correlation with carotid plaque histology.

5.5 Conclusions

Radiomic features combined with machine learning classifiers are able to differentiate asymptomatic, culprit and non-culprit carotid arteries in multi-class feature-based classification. Radiomic features alone outperformed the predictive power of FDG PET and carotid calcification. Estimates of performance by five-fold cross validation and bootstrapping were similar, with cross validation producing slightly more optimistic estimates. When using radiomic features extracted from CTA images with no pre-processing, carotid calcification provided no additional benefit to using radiomic features alone as predictors. When using radiomic features extracted from CTA images with resegmentation [0-200] HU, carotid calcification did provide additional information. For multi-class classification, resegmentation did not provide great benefit to the predictive power of the model. Carotid CTA radiomic features do relate to markers of inflammation, defined both by PET and immunohistochemistry, suggesting they may reflect important pathophysiological processes in atherosclerosis.

Chapter 6: Deep Learning for Carotid Artery Classification with Interpretability

Chapter summary:

This chapter focuses on the application of deep learning to the carotid CTA imaging dataset, investigating the use of transfer learning to overcome the limited dataset, as well as the use of 3D convolutional neural networks (CNN) for multi-class classification. In order to enhance the interpretability of the deep learning model's decisions, we use the Grad-CAM method to visualise and interpret the imaging features that the CNN is using to influence the classification decisions and we implement some deep learning stress testing to investigate how deep learning models can change their classification decisions.

6.1 Background and Rationale

Radiomics approaches use image-derived metrics to mine the information contained within a medical image. However, this feature-based approach may be limited by the extent of human expert knowledge in terms of conceptualising and developing relevant radiomic features to extract. Deep learning takes this responsibility away from us and opens up the opportunity to discover new knowledge. Neural networks, which form the core of deep learning, have the capability of learning any mapping function from the input data to the desired output data according to the universal approximation theorem[232]. However, the risk with such high capacity models is that they may fail to learn relevant features from the input data or learn relationships between the input data and the output labels that are not clinically relevant.

6.1.1 Aims and Objectives

The specific objectives for this chapter are as follows:

- To investigate the feasibility of image-based deep learning (also known as end-to-end training) for carotid artery classification in both binary and multi-class classification with transfer learning.
- To investigate the feasibility of using 3D convolutional neural networks (CNNs) for carotid artery classification.
- To compare the predictive performance of end-to-end deep learning with radiomics feature-based machine learning for carotid artery classification.
- To enhance the interpretability of deep learning algorithms using visualisations and stress testing.

6.1.2 Hypotheses

- Deep learning will out-perform radiomic feature-based machine learning in discriminating carotid artery type.
- 3D CNNs may be difficult to train with the limited data available and lack of pre-trained models for transfer learning implementation.
- The deep learning models will be susceptible to stress testing if they have overfitted to the training data.

6.2 Materials and Methods

6.2.1 Carotid CTA Imaging Dataset

The carotid imaging dataset, as mentioned in previous chapters (see Chapter 2, section 2.2.1), was used in this deep learning project to allow for direct comparison with the feature-based machine learning approach using radiomic features as predictors. A description of the acquisition of the carotid CT angiography scans can be found in Chapter 2, section 2.2.1. This dataset consisted of 132 carotid arteries: 82 carotid arteries were from symptomatic patients who previously had a stroke or TIA, 41 of which were culprit carotid arteries and the other 41 were non-culprit carotid arteries. There are 50 carotid arteries from asymptomatic patients who have not experienced any previous cerebrovascular events.

6.2.2 Analysis Tools

The programming language Python version 3.6.7 was used for the deep learning analysis in this chapter. The code was executed in a Jupyter Notebook environment, a platform that allows Python code to be run and for the code to be edited interactively. The binary classification deep learning tasks were run on a MacBook Pro (15-inch, 2017, 2.8 GHz Quad-Core Intel Core i7 processor, 16 GB Memory) laptop, and the multi-class classification tasks were run on a TensorBook Lambda laptop (Intel Core i7-8750H CPU processor, 32 GB Memory).

The project made use of several key Python libraries and packages, detailed in Table 6.99.

Table 6.99 Python libraries and packages used in this project

Purpose	***Packages and Libraries***	***Further Details***
Numerical computations	NumPy	NumPy stands for numerical Python and allows for mathematical operations to be performed on arrays (a list that only stores numeric values).
Image manipulations	Pillow (PIL), Scikit-image, OpenCV, SimpleITK	These libraries provide functions for image processing including image filtering (such as blurring), image cropping and geometric transformations (such as rotation)
Visualisations	Matplotlib	This library provides functions for plotting graphs and other visualisations.
Machine Learning and Deep Learning	TensorFlow, Keras, Keras-vis, Scikit-Learn	TensorFlow is a Python library created by Google that can be used to create deep learning models directly, or with wrapper libraries (e.g. Keras) which simplifies the process. In Keras, each line of code creates one layer of a network and there is a selection of state-of-the-art architectures and pre-trained weights that can be used for transfer learning.
Model persistence	H5Py, pickle, JSON	The h5py package allows for the storage of large amounts of numerical data. The pickle module serialises objects (converts an object in memory to a byte stream) so that they can be saved to a file and loaded into a program again later on. JSON is a standardised and programming-language independent library for the storage of data in a human-readable format.

6.2.3 Data Preparation

In end-to-end training where the input data is fed directly to the deep learning algorithm for feature extraction and classification, each pixel in the image represents a dimension of the data. Therefore, for a single CT angiography slice that has a matrix size of 512x512 pixels, this results in high-dimensional data of 262,144 dimensions. If a 3D volume is analysed by stacking multiple axial slices together (e.g. in the same manner that was used for multi-slice radiomic feature extraction using 14 consecutive slices of the carotid artery), this results in an input with a dimension of 3,670,016 (512x512x14). This high-dimensional input presents a challenge for the deep learning model to learn suitable representations from the data as the search space is vast and the available data is limited.

The axial CTA slice is therefore cropped to a 30x30 image patch centred around the carotid artery to direct the deep learning model towards the relevant anatomy within the image, where there might be a useful signal to identify the carotid artery status, i.e. asymptomatic, culprit or non-culprit carotid artery. Image patch sizes of 30x30 have been used in other deep learning applications, for example in the classification of tuberculosis from lung CT scans[233]. In addition, reducing the size of the input makes it faster to train the model and increases the possible batch sizes used as more image patches can be loaded into RAM (random access memory of the computer). The width and height of the image patches were maintained in a 1:1 ratio because convolutional neural networks typically expect a square-shaped input image.

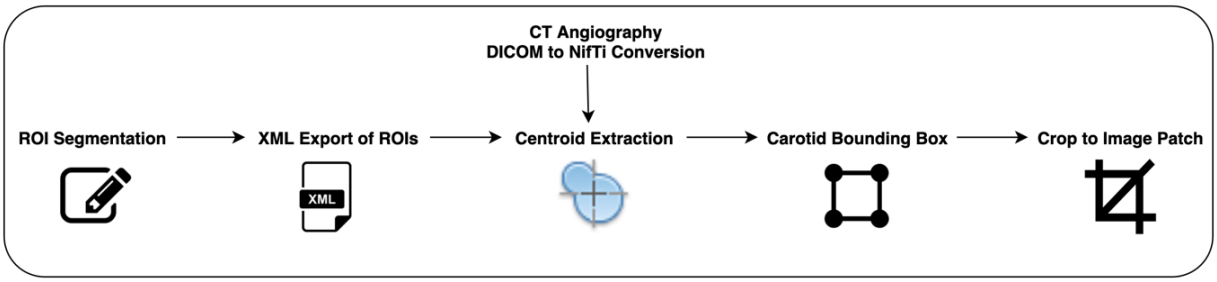


Figure 6.84 Carotid Image Patch Generation Workflow

The creation of the carotid image patches extended from the ROIs drawn around the carotid arteries in TexRAD, discussed in Chapter 2, section 2.2.4. The workflow for carotid image patch generation is shown in Figure 6.84 and consisted of seven steps described below.

Carotid Image Patch Generation Workflow:

1. Carotid arteries were manually segmented in TexRAD to create ROIs (described in Chapter 2, section 2.2.4).
2. ROIs were saved and exported as XML files (see Figure 6.85) which stores the x, y coordinates of the ROIs.
3. The CTA axial slice DICOM format was converted to NifTI files (described in Chapter 3, section 3.2.2.3).
4. The XML files and corresponding NifTI file of the CTA axial slice were loaded into a Jupyter notebook.
5. The Python module *XMLTree* extracted the x,y coordinates from the XML file.
6. The Python module regionprops extracted properties from the XML ROI: specifically, the centroid of the ROI.
7. This information was used to create a 30x30 bounding box around the carotid artery and the axial slice was cropped to this carotid image patch which was saved as a .npy file for use in deep learning classification, see Figure 6.86.

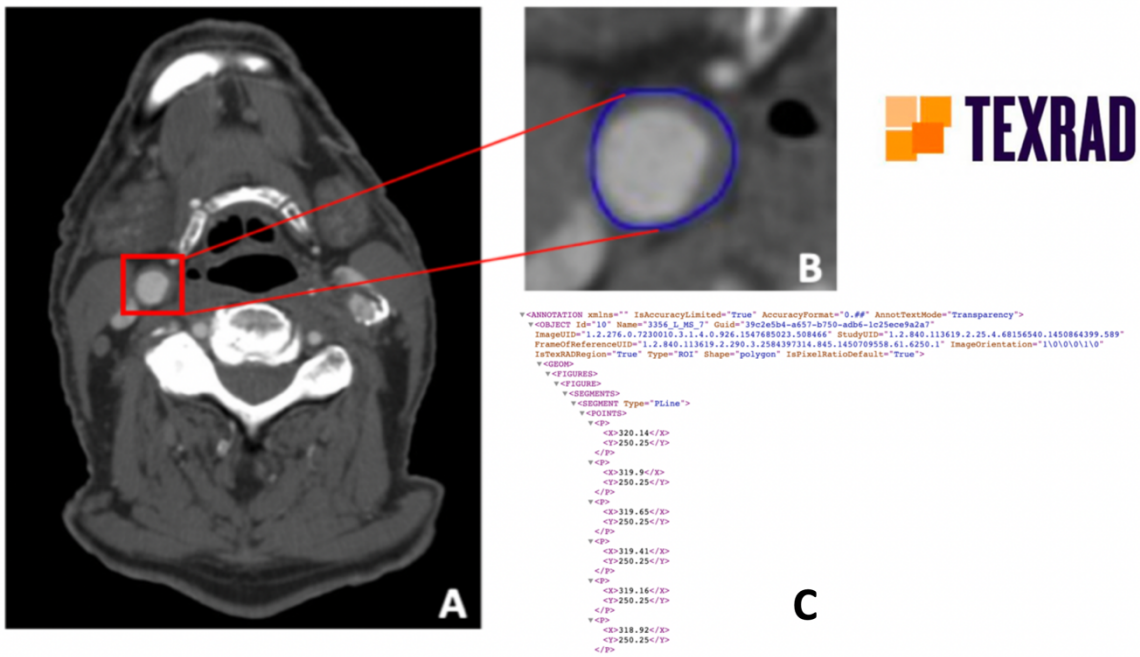


Figure 6.85 Carotid Artery Segmentation and exporting ROIs as XML files

```
regions=regionprops(mask)
```

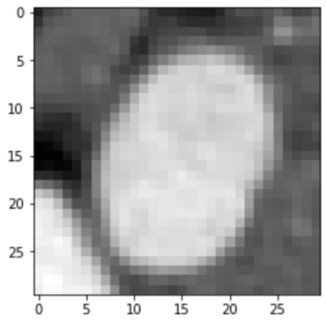
```
for props in regions:
    y0, x0 = props.centroid
```

```
min_r = int(y0-15)
max_r = int(y0+15)
min_c = int(x0-15)
max_c = int(x0+15)
```

```
square = image_arr[:, min_r:max_r, min_c:max_c]
```

```
1 plt.imshow(square[0], cmap='gray')
```

```
<matplotlib.image.AxesImage at 0x1c27a87320>
```




```
1 square.shape
```

```
(1, 30, 30)
```

Figure 6.86 Creation of a 30 x 30 carotid image patch

6.2.4 Data Pre-processing

6.2.4.1 Rescaling: Min-max normalisation

The aim of deep learning models is to learn a mapping from input variables (e.g. carotid image patch) to an output variable (e.g. the carotid artery type: 'culprit', 'non-culprit', or 'asymptomatic'). Differences in the scales of the input variables can result in a slow or unstable learning process. Scaling the input variables is therefore an important data pre-processing consideration.

In this study, the carotid image patches were rescaled from the original greyscale range to a range of -1 to 1, using the following formula:

$$x' = 2\frac{x - \min(x)}{\max(x) - \min(x)} - 1$$

Equation 6.8 Rescaling formula for range -1 to 1

where x is an original value and x' is the rescaled value[234].

This process is known as min-max normalisation and is used to reduce the sensitivity of deep learning models to differences in the scales of input variables.

6.2.4.2 Conversion from greyscale to RGB image for transfer learning

Transfer learning refers to the process of applying a model trained on a different dataset (such as colour images of cars) to a different domain (e.g. medical images). In Keras, there are several pre-trained models available whereby the model architecture and the model weights were previously learnt when they were trained on the large ImageNet dataset[145]. In order to use these pre-trained models for our own use case, the input images should be pre-processed in the same way as those that the model were originally trained on. This means that the carotid image patches, which are greyscale medical images, need to be converted into colour images with RGB (red, green and blue) colour channels. The minimum size for these input images were 48x48 pixels. Therefore, when transfer learning was implemented, the carotid image patches underwent further pre-processing which involved using *OpenCV* to resize the image array from 30x30 to 48x48 via cubic B-spline interpolation and replicating and concatenating the image array three times to produce 3 colour channels. In this way, the greyscale carotid image patches acquired the same dimensions as ImageNet images.

6.2.4.3 Preparing data for 3D Convolutional Neural Networks

As described in Chapter 2, section 2.2.4, ROIs were drawn to delineate the carotid artery on 14 consecutive slices in radiomic multi-slice analysis. In this study, when investigating the feasibility of using 3D CNNs for deep learning, we could make use of the ROIs drawn in multi-slice analysis. Each ROI was converted into a carotid image patch (as described in section 6.2.3) and for each carotid artery, the corresponding 14 carotid image patches which had been saved as .npy files were vertically stacked to form a 3D volume of dimensions 14 x 30 x 30 that was then saved as a single .npy 3D volume that could be used for training a 3D CNN.

6.2.5 Convolutional Neural Networks

The deep learning models (i.e. the convolutional neural networks) were built and trained using the Python library *Keras* with a TensorFlow backend.

6.2.5.1 Transfer Learning with 2D CNNs

6.2.5.1.1 Modified VGG16 Model Architecture

Since the carotid imaging dataset is relatively small compared to datasets normally available for deep learning and computer vision purposes, we investigated the feasibility of leveraging transfer learning to repurpose models that were previously trained on much larger datasets. There were several models available from the Keras library that were pre-trained on the ImageNet Large Scale Visual Recognition Challenge collection of images (approximately 1.2 million images of everyday objects belonging to 1000 different classes)[145]. The convolutional neural network architectures considered, but were not limited to, VGG16, Xception and ResNet50. In this study, the VGG16[235] architecture was used due to its relatively simple architecture compared with the more recent models such as ResNet50 in accordance with the principle of Occam's Razor[236].

The VGG16 model architecture was loaded into the Jupyter notebook from the Keras library and was initialised with its learned weights when it had been trained on the ImageNet challenge dataset. However, the final fully connected layers that had 1000 nodes (also known as neurons) to output predictions for the 1000 classes within the competition were removed for this study. This is because we would not be predicting 1000 classes of everyday objects, but rather this study focused on (1) a binary classification problem of either: 'culprit' versus 'non-culprit' carotid artery or 'symptomatic' or 'asymptomatic' patient or (2) a multi-class classification problem with only 3 possible outcomes: 'culprit', 'non-culprit' or 'asymptomatic' carotid artery. In order to adapt the model to our carotid artery use-case, the model architecture was modified, as illustrated in Figure 6.87. Briefly, the original VGG16 model architecture was used as the base model, initialised with ImageNet weights. Subsequently, the top fully connected layers (also known as dense layers) were removed and replaced with a global average pooling layer, a fully connected layer, batch normalisation, dropout and a final fully connected layer (with the softmax activation function) and two or three neurons depending on the number of carotid artery classes to be predicted.

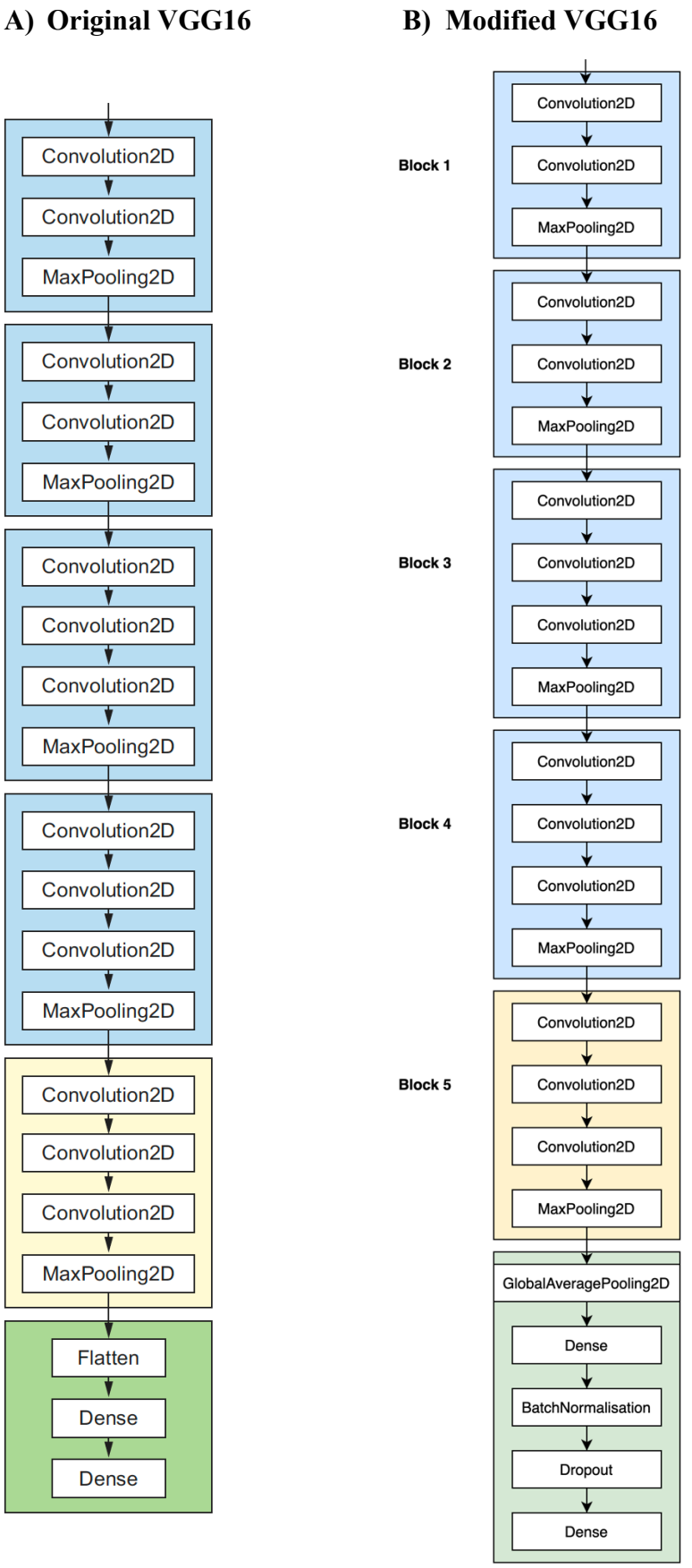


Figure 6.4A is reproduced from Deep Learning with Python[237] by François Chollet (creator of Keras) and represents the original VGG16 architecture without any modifications.

Figure 6.4B is a diagram of the modified VGG16 architecture which involved removing the original fully connected layers (represented in Figure 6A as the green block) and replacing this with a global average pooling layer, a fully connected layer, batch normalisation, dropout and the final fully connected layer with the number of nodes that corresponded to the number of classes to be predicted.

Figure 6.87 Original versus Modified VGG16 Model Architecture for Transfer Learning

The global average pooling layer calculates the average value for each feature map from the previous layers and is a down sampling method used to reduce the complexity of a model, reduce computational cost and risk of overfitting[238]. In addition, it has been found to improve the ability of CNNs to learn discriminative regions of images related to class-specific decisions[239].

Batch normalisation and dropout are regularisation techniques which aim to reduce the likelihood of overfitting on the training dataset. Batch normalisation standardises the outputs of the previous hidden layer for the subsequent input layer for every mini-batch. This stabilises and speeds up the training process of deep neural networks[240] and is used by default in many deep learning applications[241].

Dropout regularisation works by randomly selecting neurons to be 'dropped-out' during training at a given probability (this is usually set at 20-50%), in which their value is set to zero. Therefore, their contribution to downstream neurons is temporarily removed, ensuring that the deep learning model as a whole does not become over-reliant on the specific weights of particular neurons. Dropout encourages the model to learn multiple independent representations of the data with the aim of improving the model's ability to generalise on new datasets[242]. In this study, the dropout probability was set to the commonly used value of 0.2 which randomly disables 20% of the hidden layer units. The value of the dropout probability had to be decided by the operator and is referred to as a hyperparameter.

6.2.5.1.2 Hyperparameter Choices

Hyperparameters refer to the set of parameters which need to be pre-specified prior to training the deep learning algorithm[243]. There are several hyperparameters, all of which can impact the deep learning model's performance. However, it is difficult to use a data-driven approach to find the set of optimal hyperparameters as certain choices are discrete rather than continuous (unlike learning weights for the neural networks) and it is not easy to use gradient descent to search the vast hyperparameter space. Hyperparameter tuning methods currently include random search whereby different values for hyperparameters are chosen and evaluated on the validation set, or grid search, where a list of different values for the hyperparameters are iteratively evaluated on the validation set. Hyperparameter optimisation can therefore lead to overfitting of the validation dataset. In this study, our focus was not on finding the optimal set

of hyperparameters via hyperparameter tuning, but rather on the feasibility of using deep learning for carotid artery classification. As a result, we did not perform automatic hyperparameter optimisation in this study to avoid overfitting on the validation dataset. In Table 6.100, we consider the different hyperparameters associated with deep learning and specify the values that were used in this study.

Table 6.100 Hyperparameter definitions and decisions

Hyperparameter	Description	Values used in this study
Activation function	Each neuron/node has an activation function which is applied to its input to generate the node's output and often helps to model non-linear functions[137]. Different activation functions exist such as sigmoid, tanh, softmax or the rectified linear activation function (ReLU) activation functions.	For all layers of the CNN (except the final dense layer), the ReLU activation function was used, which outputs zero for negative inputs and the identity for positive inputs. ReLU is the most popular activation function in deep learning and has been shown to work well (converge more quickly) empirically[244]. The softmax activation function[237] was used in the final fully connected layer to convert the layer's inputs into a probability distribution between 0 and 1.
Objective function (also 'cost' or 'loss' function)	The evaluation metric that measures the gap between the model's prediction and the ground truth. This gap (or difference) is the error that we want to minimise by optimising (i.e. adjusting) the weights and biases of the deep learning model with the backpropagation algorithm that aims to move the weights in the right direction.	For binary classification tasks, we used binary cross entropy loss. Cross entropy originates from Information Theory and measures the distance between probability distributions, or in this case, between the ground truth distribution and the model's current predictions. For multi-class tasks, categorical cross entropy loss was used[237].

Learning rate	The learning rate determines the magnitude of how much the weights and biases are adjusted after each mini-batch (see Batch Size below). If the learning rate is too large, the weight updates could lead to an unstable learning process that does not find a suitable solution. Conversely, if the learning rate is too small, the learning process would be too slow.	Learning rate values typically range from 1×10^{-7} to 1. The initial learning rate used in this study was 0.0001, which is on the lower end as we wanted to limit the magnitude of modifications made to the weights in the transfer learning model to prevent large updates from harming useful representations already captured by the ImageNet weights[237]. Furthermore, this initial learning rate has been shown to be a good initialisation trick[245]. A learning rate decay of was also used to slowly decrease the learning rate over the training process as this was shown to improve generalisability[246] empirically.
Optimiser	Optimisers refer to the algorithms used to update the weights and learning rate in order to minimise the model's error. Modern optimisation algorithms are essentially extensions of the stochastic gradient descent algorithm which calculates the gradient of the error and updates weights in the direction opposite to this gradient after every training sample[247].	RMSProp, one of the most popular optimisers in deep learning, was used. It builds upon the older stochastic gradient descent optimisation algorithm and is an adaptive learning rate method[248]. This involves monitoring the performance of the model on the training dataset and adjusting the learning rate in response to this.

Batch Size	The batch size determines how many data samples are propagated through the model before the weights are adjusted. If all the data is used before updating the weights, this is referred to as a batch. If a subset of the data is used each time, this is referred to as a mini-batch.	A batch size of 8 was used as it is recommended to use a batch size in the power of 2 for better performance[249] and there would be sufficient RAM to handle this batch size.

6.2.5.1.3 Training the VGG16 Layers

As demonstrated in Figure 6.87, the VGG16 architecture consists of 5 blocks of convolutional layers which act as the feature extractors of the deep learning model. The final block (coloured in green in Figure 6.87) of the model acts as the classifier which maps the features from the feature maps of the previous layers to the different classes. The shallower layers of the model (e.g. block 1) when initialised with the pre-trained ImageNet weights represent features that are low-level and generic such as edges, lines and textures. These features are useful for all types of computer vision tasks from real world images to medical images. However, the representations learnt in the deeper layers of the model (e.g. block 5) would be more specific to the classes in ImageNet and represent more abstract concepts such as 'dog ears'[237] which are not so relevant for the carotid artery use case.

Therefore, in training the deep learning model using transfer learning, the convolutional layers of VGG16 up until block 5 convolutional layer 1 were frozen. This meant that the weights of all the layers before block 5 would not be updated as the model went through the carotid imaging dataset during training. This preserved the useful generic representations learnt by the pretrained ImageNet weights. The layers from block 5 onwards including the layers that were added to create the modified VGG16 architecture in Figure 6.87B, were set to trainable. This permitted the model to be fine-tuned for this carotid artery use-case, as the weights for these layers would be adjusted in relation to the carotid artery classification tasks.

6.2.5.1.3 2D Transfer Learning Sample Sizes in Binary and Multi-class Classification

When using transfer learning with the 2D CNN, three levels of analysis were conducted: (1) binary classification at the artery level of culprit versus non-culprit carotid artery, (2) binary classification at the patient level of asymptomatic versus symptomatic patient and (3) multi-class classification at the artery level of asymptomatic versus culprit versus non-culprit carotid artery. Details of the dataset sample sizes used in each level of analysis is provided in Table 6.101.

Table 6.101 Dataset sample sizes in binary and multi-class classification

Analysis	**Dataset**
Binary classification: artery level Culprit versus non-culprit carotid arteries	Used carotid patches of the carotid bifurcation only (derived from ROIs drawn in single-slice analysis in Chapter 2, section 2.2.4.3). There were 82 carotid arteries in total from 41 symptomatic patients, resulting in 41 culprit and 41 non-culprit carotid arteries. The dataset was split such that 75% was used for training and 25% was used for validation. This translated to 61 carotid bifurcation image patches of size 30x30 being used for training and 21 used for validation. Please note for this experiment, the results from the VGG16 model (transfer learning) were compared to a ResNet18 architecture and a simpler LeNet architecture that were trained from scratch to assess the value of transfer learning in improving predictive ability.
Binary classification: patient level Asymptomatic versus symptomatic patients (more specifically identifying whether the carotid artery	To expand the available dataset, each ROI drawn in multi-slice analysis (see Chapter 2, section 2.2.4.4) was converted into an individual CTA carotid image patch. As there were 14 ROIs drawn per carotid artery in multi-slice analysis, this resulted in 14 carotid image patches per carotid artery. There were 82 carotid arteries from symptomatic patients and therefore a total of 1148 carotid image patches of symptomatic carotid arteries (14x82). There were 50 asymptomatic carotid arteries, giving rise to a total of 700

belonged to an asymptomatic patient or a symptomatic patient)	carotid image patches of asymptomatic carotid arteries. The total dataset therefore consisted of 1848 carotid image patches. The dataset was split such that 75% was used for training (1386 carotid patches) and 25% for validation (462 carotid patches).
Multi-class classification: artery level Asymptomatic versus culprit versus non-culprit carotid arteries	To follow the dataset set up used in Chapter 5's multi-class classification whereby 120 carotid arteries were used for training and 12 where used for external validation, this level of analysis used the carotid image patches from the same 120 carotid arteries for the train-validation split and the carotid image patches from the other 12 carotid arteries were set aside for independent testing. The entire dataset available for training was therefore 120x14 = 1680 carotid image patches. Following training-validation split, 1260 carotid image patches were used for training and 420 were used for validation. For external validation, the carotid image patches from the 12 carotid arteries (12x14 = 168 carotid image patches) that were previously used for external validation of the feature-based radiomics machine learning approach in Chapter 5, were used. This was to ensure comparability between the multi-class classification tasks of Chapter 5 (a feature-based radiomics ML approach) and the deep learning approaches of Chapter 6.

Prior to train-validation split, the dataset was always randomly shuffled so that the model did not simply learn the order that the dataset was fed into the model, using a random seed of 42. Furthermore, the train-validation split was always stratified by class label.

6.2.5.1.4 Data Augmentation

A Keras generator was used to dynamically augment the data from the training dataset in real-time as it was inputted to the deep learning model. Data augmentation helps to make models invariant to small variations such as slight rotations. The training images could be randomly rotated up to 8 degrees, zoomed in by up to a factor of 0.1 and randomly translated vertically or horizontally by a factor of 0.1. No flipping was applied in order to preserve information about the left-right axis of the human body within the input image patches. No data augmentation was applied to the external validation data.

6.2.5.1.5 Early Stopping and Model Checkpointing

The training dataset was used to update the weights and biases of the convolutional neural network. The validation dataset was not used to update the parameters of the model, but to monitor whether the model was overfitting (over-learning) the training data. This could be identified by plotting the information from the model training history which contains information about the loss (the error) and the accuracy on both the training set and the validation set respectively. A model that was overfitting on the training data would have a downward sloping loss curve on the training set with an increasing accuracy that would start to plateau at high accuracy, whereas the loss curve would be increasing in the validation data with fluctuating accuracy.

In order to prevent overfitting on the training dataset, early stopping was used which is a method that achieves similar results to regularisation. This method involves terminating the training process earlier than a predefined number of epochs (e.g. 100 epochs), whereby a single epoch refers to when the whole dataset has been forward- and backpropagated through the network. In this study, the predefined number of epochs was set to 100. Early stopping would monitor the accuracy on the validation dataset and stop the training process when there was no further improvement after 10 consecutive epochs (the default setting in Keras). Early stopping is one of the functions available from Keras' callbacks, a collection of functions that extend the functionality for training neural networks.

Another callback known as checkpointing was also used. Model checkpointing monitors a specified metric such as validation accuracy during the training process and enables snapshots

of the model weights to be saved. In this study, the checkpointing callback monitored the validation accuracy and saved the model weights each time there was an observed improvement. The h5py library was used in conjunction to enable the model architecture and model parameters to be saved as a HDF5 file and ensure model persistence.

The combination of early stopping and model checkpointing permitted the model with the highest validation accuracy to be saved before the model was greatly overfitting on the training data.

6.2.5.2 Deep learning with a 3D Convolutional Neural Network

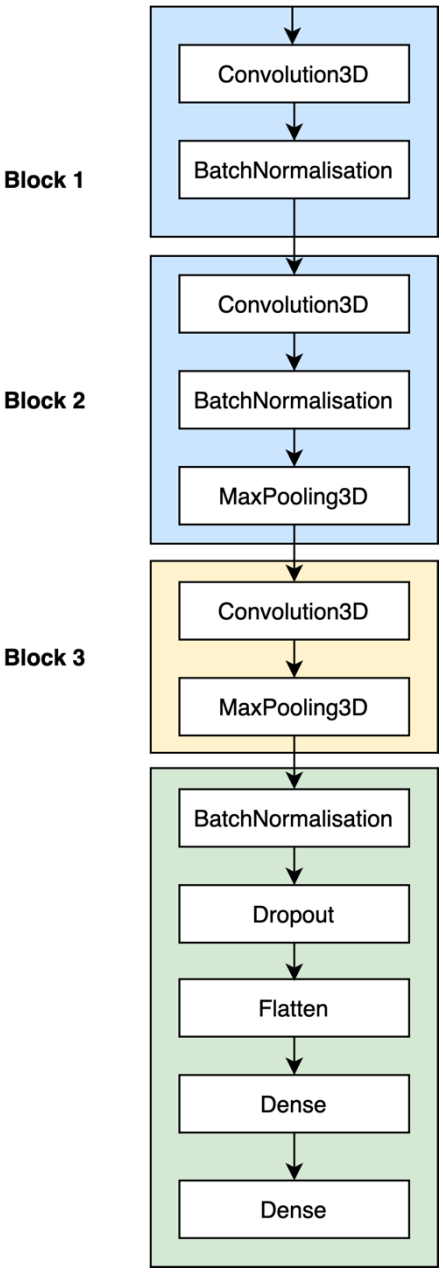
Since VGG16 was originally trained on 2D images, the pretrained ImageNet weights available from Keras only work for the architectures that deal with 2D images. In this study, we also wanted to explore the use of a 3D volume (as in Chapter 3 for multi-slice radiomics analysis with PyRadiomics) whereby the input consisted of 14 axial CTA carotid patches concatenated together. The dimension of each 3D input was therefore 14x30x30 and transfer learning could not be used as 3D pretrained weights are not yet available in Keras. Therefore, a simple 3D CNN model was built using Keras that would need to be trained from scratch rather than using pretrained weights. A summary of the model architecture is provided in Figure 6.88. This 3D CNN was used for the multi-class classification of asymptomatic, culprit and non-culprit carotid arteries. As such, the dataset consisted of 40 asymptomatic, 40 culprit and 40 non-culprit carotid 3D volumes which was split 75% for training and 25% for internal validation (stratified by class label). This resulted in a total of 90 carotid 3D volumes used for training and 30 carotid 3D volumes used for validation. In addition, there were 12 separate carotid 3D volumes held out for external validation (as has been the case for other multi-class classification tasks, see Table 6.101).

The dataset was shuffled prior to input to the model for training and min-max normalisation was applied to the 3D input as previously described in section 6.2.4.1. A batch size of 3 was used to accommodate for restrictions of RAM. The activation function used for all layers of the 3D CNN, except the final fully connected layer, was ReLU. The softmax activation function was used for the final fully connected layer to output probabilities. The initial random weights of the model prior to training were set by the Keras default, which was the Glorot uniform initialiser (also known as the Xavier uniform initialiser)[250]. The categorical

crossentropy loss was used along with the RMSprop optimizer with an initial learning rate of 0.0001 and a learning rate decay of $1x10^{-6}$. The number of epochs the model would be trained for was prespecified as 100, but early stopping and model checkpointing was used as described in section 6.2.5.1.5.

```
print(model.summary())
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv3d (Conv3D)              (None, 13, 28, 28, 32)    608
_________________________________________________________________
batch_normalization (BatchNo (None, 13, 28, 28, 32)    128
_________________________________________________________________
conv3d_1 (Conv3D)            (None, 12, 26, 26, 64)    36928
_________________________________________________________________
batch_normalization_1 (Batch (None, 12, 26, 26, 64)    256
_________________________________________________________________
max_pooling3d (MaxPooling3D) (None, 6, 13, 13, 64)     0
_________________________________________________________________
conv3d_2 (Conv3D)            (None, 5, 11, 11, 128)    147584
_________________________________________________________________
max_pooling3d_1 (MaxPooling3 (None, 2, 5, 5, 128)      0
_________________________________________________________________
batch_normalization_2 (Batch (None, 2, 5, 5, 128)      512
_________________________________________________________________
dropout (Dropout)            (None, 2, 5, 5, 128)      0
_________________________________________________________________
flatten (Flatten)            (None, 6400)              0
_________________________________________________________________
dense (Dense)                (None, 64)                409664
_________________________________________________________________
dense_1 (Dense)              (None, 3)                 195
=================================================================
```



A) Architecture of 3D CNN as summarised by Keras

B) Simplified diagram of simple 3D CNN

Figure 6.88 Model Architecture of the Simple 3D CNN

The simple 3D CNN model consists of a total of 3 convolutional layers with the addition of batch normalisation, dropout and maxpooling as regularisation and downsampling techniques respectively. The final layers of the model are fully connected and map the feature maps to the classes (asymptomatic, culprit or non-culprit carotid arteries).

6.2.5.3 Interpretability and Visualisations

The European Union's General Data Protection Regulations[251] emphasised the importance of enhancing transparency in the interpretability of AI decisions and so we investigated incorporating elements of explainable AI into our workflow. This would allow us to better understand how the neural network models arrived at their predictions which could potentially reveal new knowledge, as well as act as a sanity checkpoint to ensure the model was learning relevant features. Several strategies have been developed to gain further insight into the decisions of deep learning models and some of these strategies were employed, below.

6.2.5.3.1 Grad-CAM visualisations

The Grad-CAM algorithm was developed by Selvaraju et al. in 2017 and stands for gradient-weighted class activation mapping[252]. This algorithm is useful for understanding which parts of a given image contributes to the model's classification decision for a particular class e.g. culprit carotid artery and can help to identify whether the model has learnt to 'look' in the correct locations of the image. For a given input image, the algorithm examines the gradient information flowing into a specified layer (e.g. the final convolutional layer) of the model and outputs a heatmap (a coarse localisation map) that highlights the important regions of the image for predicting a given class label. The maximally activating regions were highlighted in red and the minimally activating regions were blue.

Originally, Grad-CAM heatmaps were created with the Keras visualisation library keras-vis, which took a trained deep learning model, a layer index to specify which layer to visualise and an input image, which was propagated through the network. However, the code for the keras-vis library was no longer maintained by the time the multi-class classification experiments were conducted. To overcome this, class activation visualisations for the multi-class classification tasks used a method directly implemented with Keras, without the need for the keras-vis library. This method was devised by François Chollet (the creator of Keras) and the code can be found here: https://keras.io/examples/vision/grad_cam/.

6.2.5.3.2 Visualising CNN intermediate activations

Visualisations of the intermediate convolutional layer outputs, referred to as the intermediate activations, were examined to help us understand what the CNN was extracting from the input images. This consisted of displaying the feature maps produced by the different convolutional layers (or pooling layers in necessary) in a network when given a certain input image. Visualising CNN intermediate activations provides us with an idea of how an input is decomposed by the network and how successive layers in the CNN transforms their inputs.

6.2.5.3.3 Deep Learning Stress Testing

Various computer vision operations were applied to input images after the deep learning model was trained to assess the impact of these operations on the model's classification probabilities with respect to each given class label.

The following operations were applied, either alone or in different combinations:

1. Gaussian smoothing using the SciPy ndimage submodule for image processing function 'ndimage.gaussian' filter. This had the effect of blurring details in the image by averaging out the image noise with a local neighbourhood. The size of the local neighbourhood was specified by the sigma value provided. Different values of sigma were investigated: 0.5, 1 and 1.5.
2. Addition of random noise from a uniform distribution over [0, 1] using the NumPy function 'np.random.rand'. The level of random noise mixed with the image was controlled by alpha blending values of 0.3, 0.6 and 0.9 whereby alpha blending refers to 'the process of overlaying a foreground image with transparency over a background image'[253].
3. Vertical and horizontal flipping of the image using the Numpy functions 'np.flipud' and 'np.fliplr'.

6.3 Results

6.3.1 2D Convolutional Neural Networks (Transfer Learning)

6.3.1.1 Binary classification (artery level): culprit versus non-culprit carotid arteries

This binary classification task aimed to discriminate between culprit and non-culprit carotid arteries in symptomatic patients. This involved using the carotid image patches derived from ROIs drawn in single-slice analysis (Chapter 2, section 2.2.4.3) of the carotid bifurcation only. The training set consisted of 61 carotid bifurcation image patches and the validation set consisted of 21 image patches. Data augmentation was applied to the inputs, see Figure 6.89.

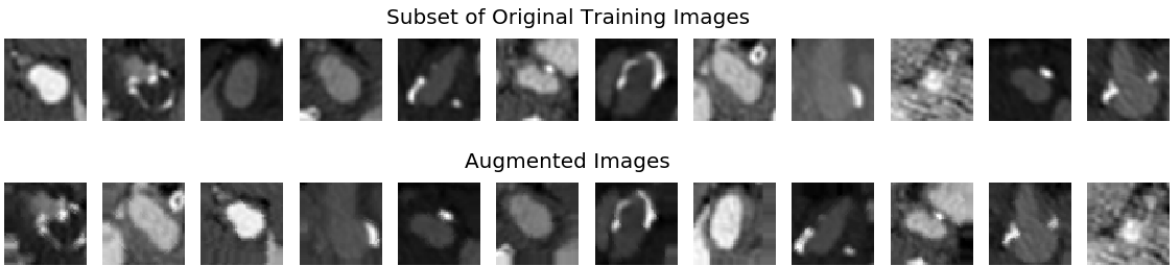
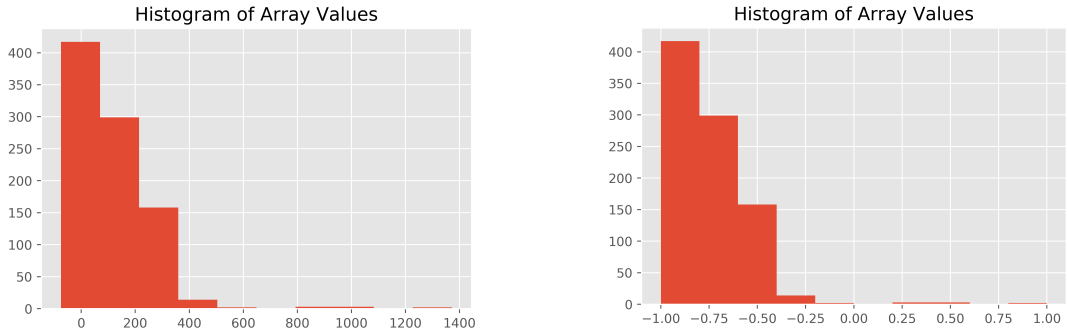


Figure 6.89 Original Training images and Images after Data Augmentation

Top: selection of training set images; bottom: selection of images with data augmentation

Additionally, the grey-values of the image patches were scaled to [-1, 1] which maintained the distribution of the HU information whilst improving the training process. This is demonstrated in the histograms of the grey-value distributions in the carotid image patches in Figure 6.90.



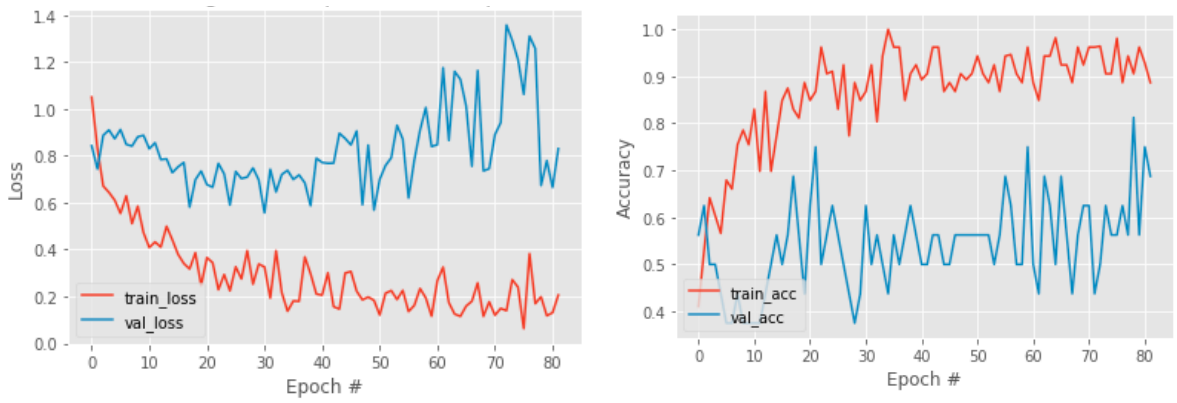
A) Grey-value distributions without normalisation

B) Grey-value distributions following min-max normalisation

Figure 6.90 Image histograms before and after min-max normalisation

Following min-max normalisation, the distribution of grey-values is the same as before. However, the magnitude and range of the values change from [-100, 1400] to [-1, 1] which speeds up the adjustments of weights during the training process and convergence of the deep learning algorithm[2].

The loss and accuracy curves for the training and validation sets are shown in Figure 6.91. The validation accuracy achieved was 71.4%, with evidence of overfitting on the small dataset as the loss curves for training and validation sets are diverging in Figure 6.91A. The incorrect predictions on this dataset are illustratrated in Figure 6.92.



A) Loss curves

B) Accuracy curves

Figure 6.91 Loss and Accuracy Curves for Training and Validation sets for Binary Classification: Culprit versus Non-culprit Carotid Arteries

[2] For the interested reader, a discussion of why feature scaling improves the speed of convergence for the deep learning algorithm, can be found here: https://datascience.stackexchange.com/questions/55656/why-does-feature-scaling-improve-the-convergence-speed-for-gradient-descent

For this experiment, the results were also compared to a ResNet18 architecture and a simpler LeNet architecture that was trained from scratch (rather than using transfer learning) using the same real-time data augmentation and hyperparameter settings. ResNet18 achieved 58% validation accuracy, whilst LeNet achieved 57% accuracy. This suggests that transfer learning did improve predictive performance.

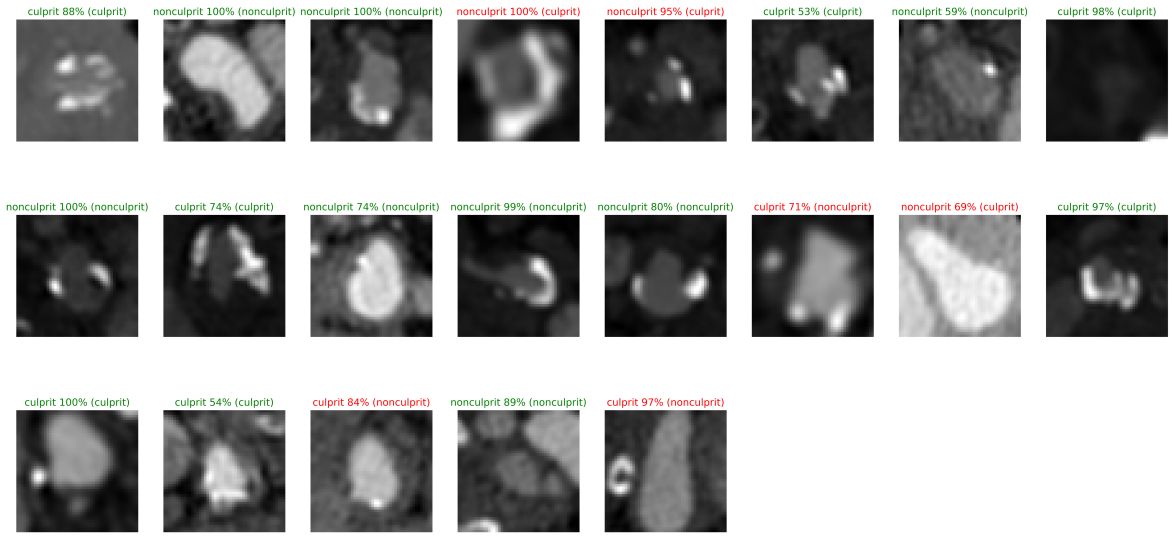
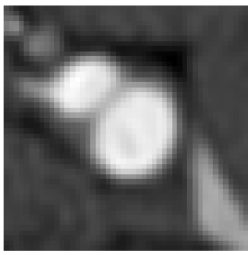


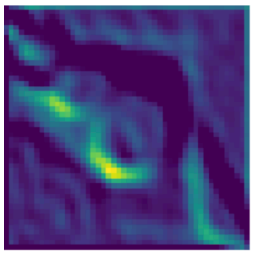
Figure 6.92 Binary classification: culprit versus non-culprit carotid arteries - correct classifications and misclassifications

Validation set images supplied to the modified VGG16 model for binary classification of culprit versus non-culprit carotid arteries. The model was not trained on these images. The model's predictions are the first label and the ground truth labels are in brackets. Green indicates the correct predictions and red indicates the incorrect predictions. The probability or 'confidence' of the model (as calculated by the softmax activation function in the fully connected layer) associated with the classification decision per image, is provided as a percentage.

For this binary classification task, visualisations of the activations of different convolutional layers of the model were produced. Figure 6.93 displays the activations of different convolutional layers that are earlier on within the network, as well as in the later blocks of the VGG16 architecture. We can see that the first convolutional layer is similar to edge detectors and was able to extract the contour of the carotid artery wall. At this stage, the activations retain most of the information present in the initial picture. However, as we progress through to later convolutional layers, the activations become more abstract and less visually interpretable as they encode higher-level concepts. Most filters are activated by the input image, but in the following layers with increasing depth, more and more filters are blank. This means the pattern encoded by the filter was not found in the input image. The features of a deep neural network become more abstract with the depth of the layer.



A) Original Carotid Image Patch



B) Feature map produced from input image

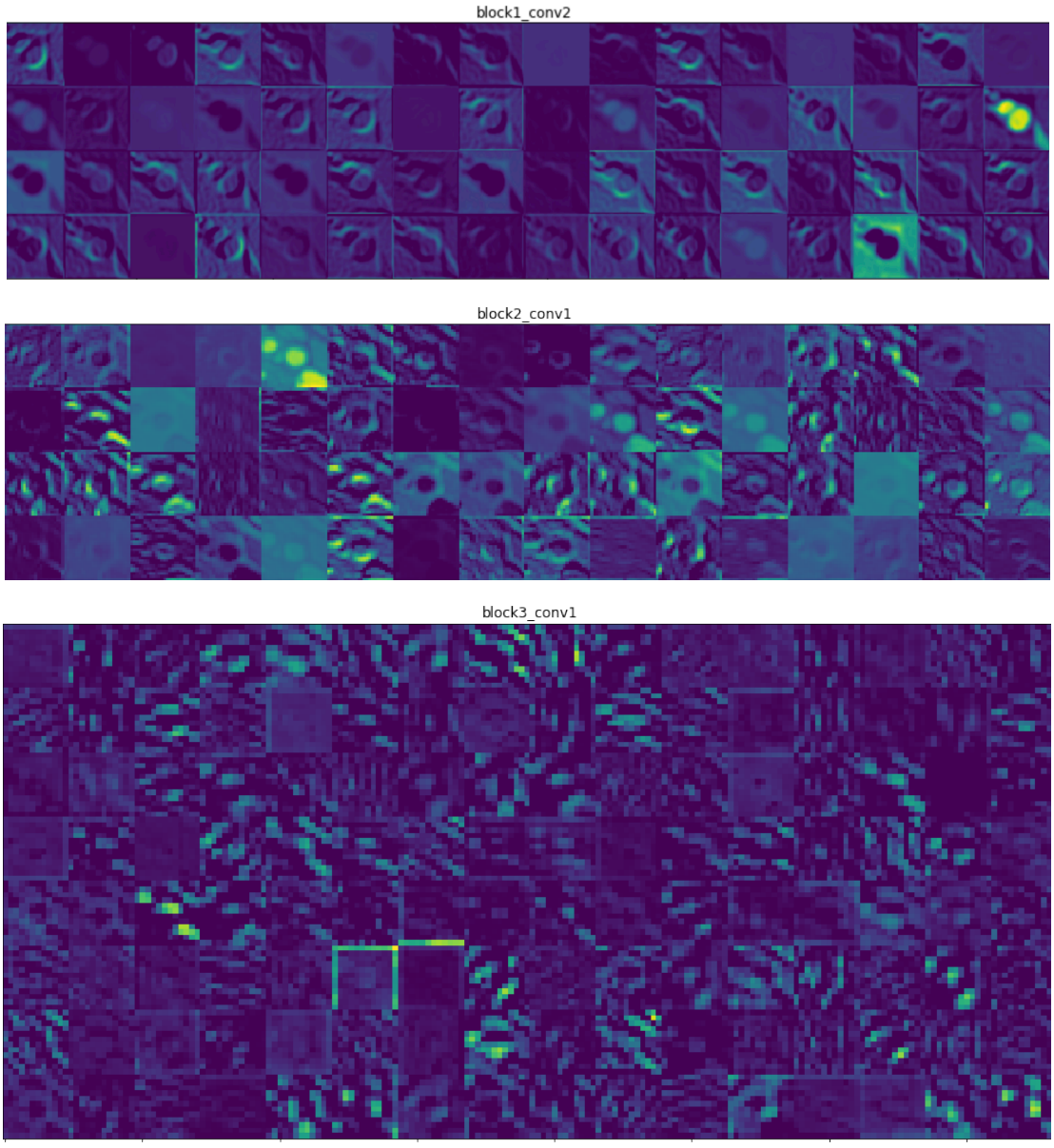


Figure 6.93 Visualisations of the activations from different layers in the CNN

These are the feature maps that are outputted by the different convolutional layers when given a certain input. The output of the layer is called its activation. Block 1 represents earlier convolutional layers, the higher the block number, the deeper the convolutional layers within the deep learning model. CNN, convolutional neural network.

6.3.1.2 Binary classification (patient level): asymptomatic versus symptomatic patient

For this binary classification task, the aim was to identify whether the carotid artery belonged to a symptomatic patient or an asymptomatic patient. In this task, the mean accuracy achieved on 462 image patches in the validation set was 92.6% and the model's correct predictions were associated with greater confidence in their classification decisions than in the incorrect predictions, shown in Figure 6.94.

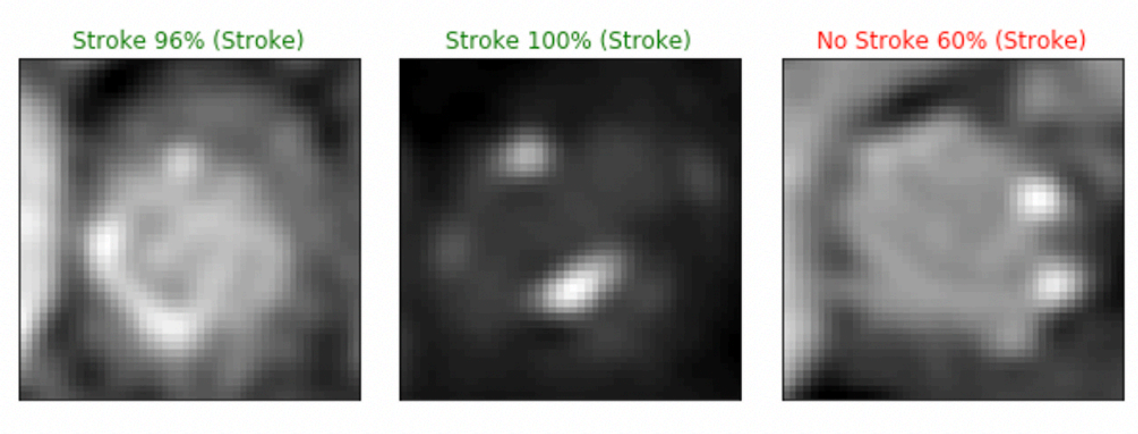
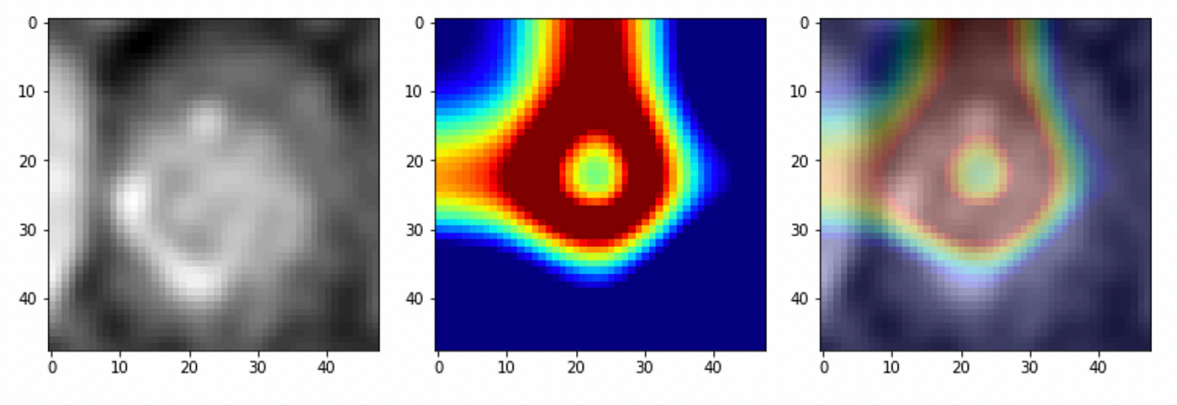


Figure 6.94 Asymptomatic versus Symptomatic Patient Validation Set Predictions

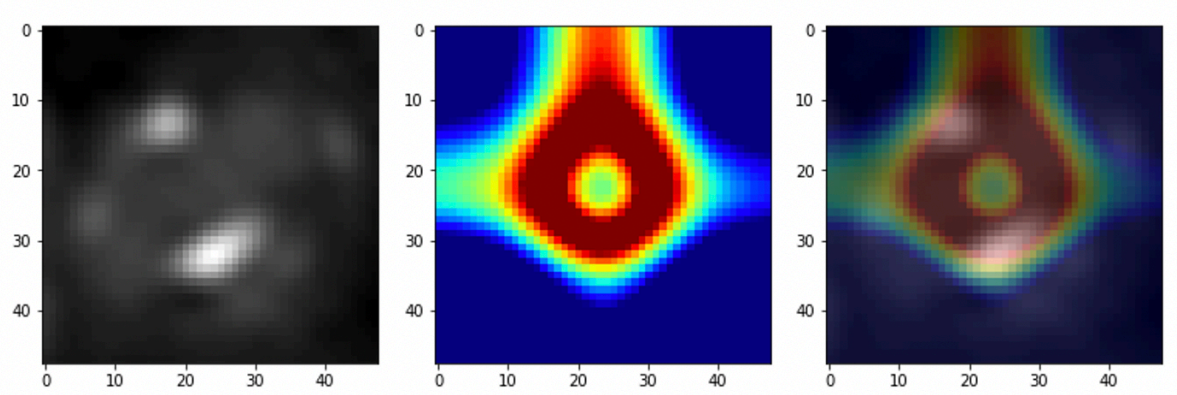
The correct predictions are highlighted in green, the incorrect prediction is highlighted in red. The probability associated with the class label prediction is given as a percentage. The ground truth label is provided in brackets.

To further explore which parts of the image patches led to the model's classification decisions, class activation heatmaps were generated using the Grad-CAM algorithm. This suggested that the model had learnt to localise the carotid artery from within the carotid image patches, as demonstrated in Figure 6.95.

A) 'Symptomatic Patient': carotid image patch from below carotid bifurcation



B) 'Symptomatic Patient': carotid image patch at the carotid bifurcation



C) 'Symptomatic Patient': carotid image patch from above carotid bifurcation

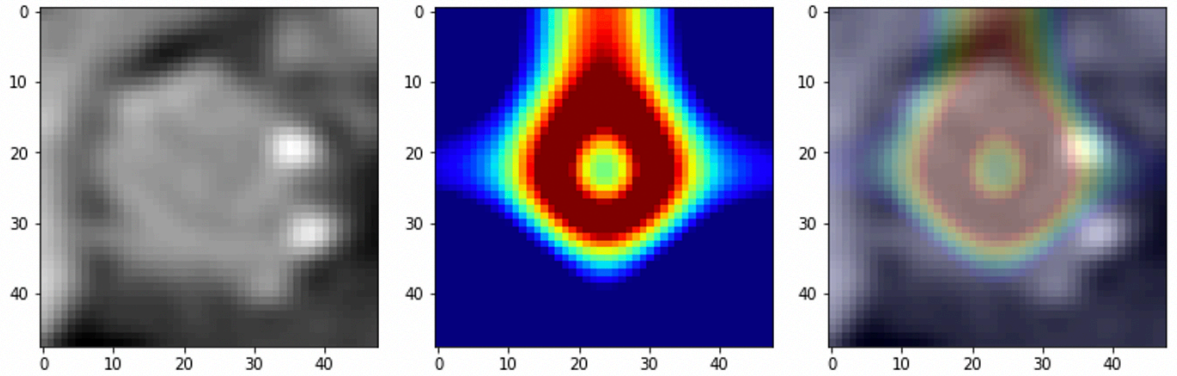


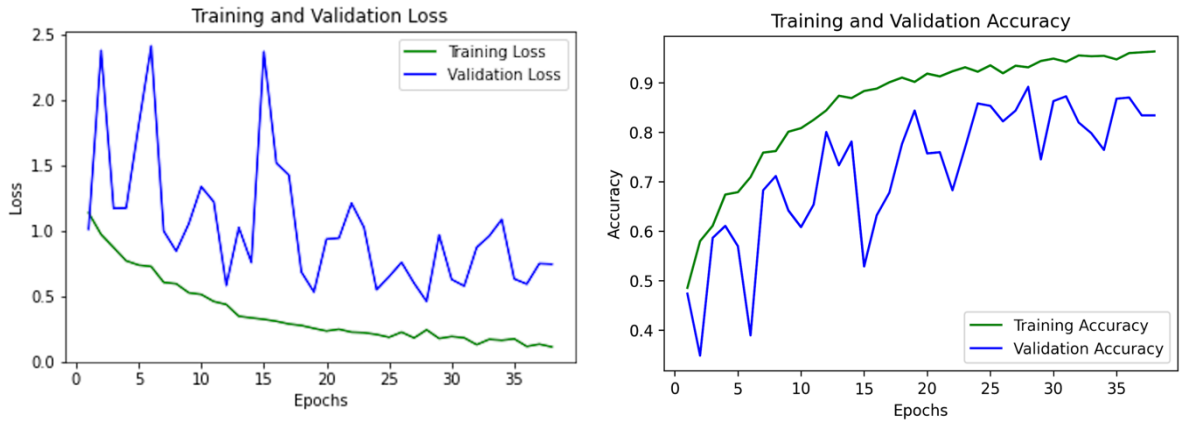
Figure 6.95 Asymptomatic versus Symptomatic Patients: Class Activation Heatmaps

The first image in each panel represents the original input carotid image patch, the second image represents the class activation heatmap where red indicates the maximally activating areas and blue the minimally activating areas. The third image provides an overlay of the heatmap over the original image. In this example, the model has localised the carotid artery within the image patch.

6.3.1.3 Multi-class classification: carotid artery level

The multi-class classification task aimed to distinguish between asymptomatic, culprit and non-culprit carotid arteries. The distribution of carotid artery types in the training-validation dataset was evenly split: 40 asymptomatic, 40 culprit and 40 non-culprit carotid arteries.

For this task, the modified VGG16 network used the categorical cross entropy loss with the RMSprop optimiser (learning rate = 0.0001, learning rate decay = $1x10^{-6}$, metrics = accuracy), a batch size of 8 and early stopping with model checkpointing, which monitored the validation accuracy and saved the best model. Figure 6.96 provides the training and validation set loss and accuracy curves.



A) Loss curves **B) Accuracy curves**

Figure 6.96 Loss and Accuracy Curves for Training and Validation Sets in Multi-class Classification

The model was able to reach 100% accuracy on the training dataset and 89% accuracy with an AUC of 0.96 on the internal validation dataset (25% of dataset, 420 carotid image patches). The confusion matrix for correct and incorrect predictions for the carotid image patches of the internal validation dataset is shown in Table 6.102.

Table 6.102 Multi-class Classification: Internal Validation Confusion Matrix

Confusion matrix	Predicted Values		
True Values	**Asx**	**CC**	**NC**
Asx	**121**	9	10
CC	2	**128**	10
NC	2	12	**126**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the CNN classifier are highlighted in red.

However, evaluation of the model on the external validation dataset only achieved a mean accuracy of 36% and an AUC of 0.46 on 168 carotid image patches (of these 14 represented culprit carotid arteries, 14 represented non-culprit carotid arteries and 140 represented asymptomatic carotid arteries).

The confusion matrix for the model's performance on the external validation dataset is shown in Table 6.103. The sensitivity and specificity respectively for each carotid artery class was as follows: Asx (37%, 54%), CC (0%, 80%) and NC (64%, 59%).

Table 6.103 Multi-class Classification: External Validation Confusion Matrix

Confusion matrix	Predicted Values		
True Values	**Asx**	**CC**	**NC**
Asx	**52**	30	58
CC	9	**0**	5
NC	4	1	**9**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the CNN classifier are highlighted in red.

GradCAM class activation heatmaps were used to identify the areas that the model was looking at for the training dataset (Figure 6.97), the internal validation set (Figure 6.98) and the external validation set (Figure 6.99).

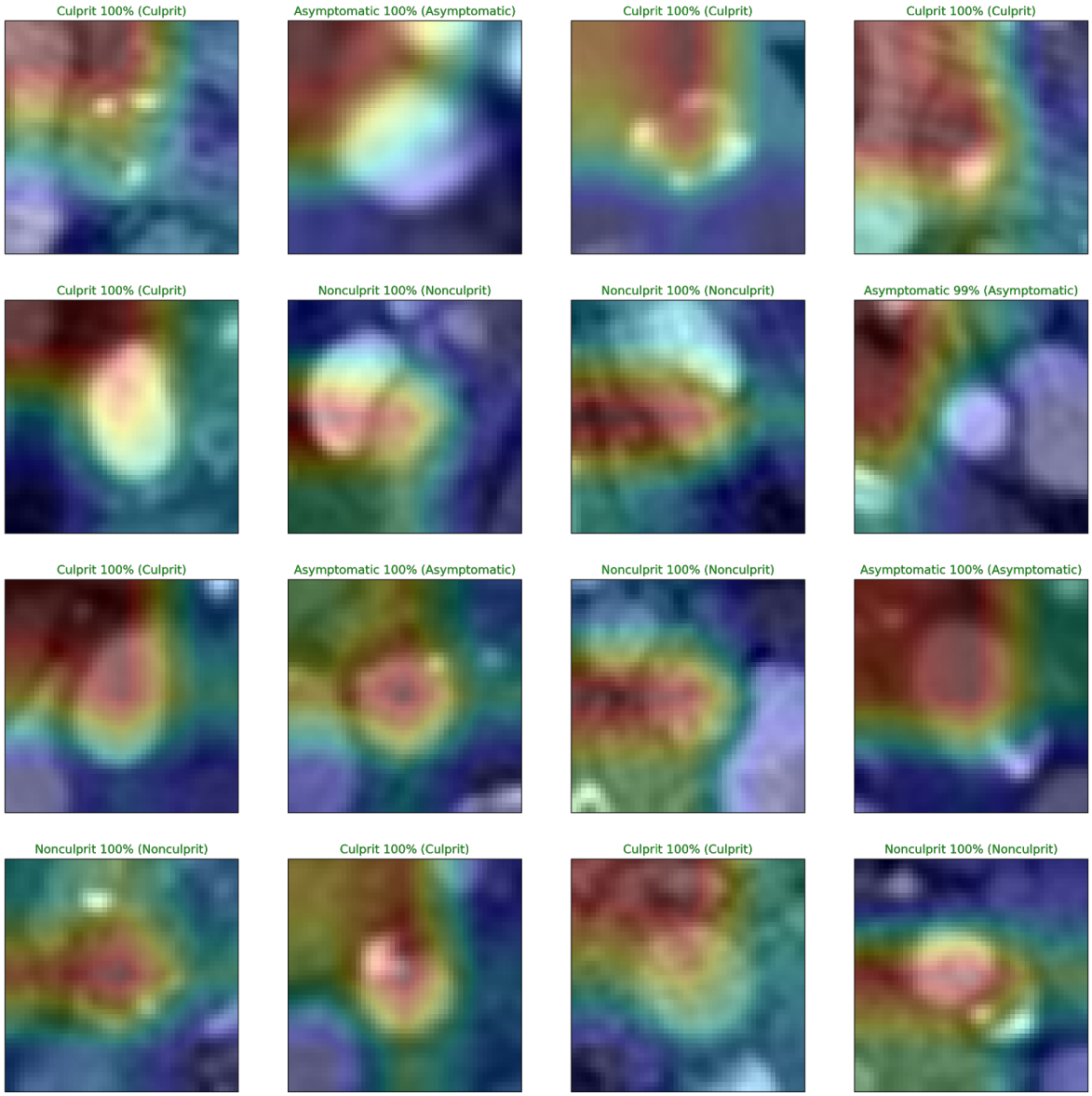


Figure 6.97 Multi-class Classification Training Data Class Activation Maps

The model's predictions are the first label and the ground truth labels are in brackets. Green indicates the correct predictions and red indicates the incorrect predictions. The probability or 'confidence' of the model (as calculated by the softmax activation function in the fully connected layer) associated with the classification decision per image, is provided as a percentage. Areas of maximal activation are highlighted in red and regions of minimal activation are highlighted in blue within the class activation heatmap.

When visually comparing the GradCAM heatmaps from the training data (Figure 6.97), internal validation data (Figure 6.98) and external validation data (Figure 6.99), there is a decrease in frequency of localising the carotid artery within the carotid image patches. This is associated with an increase in the number of incorrect predictions and a decrease in the confidence of the model in its classification predictions.

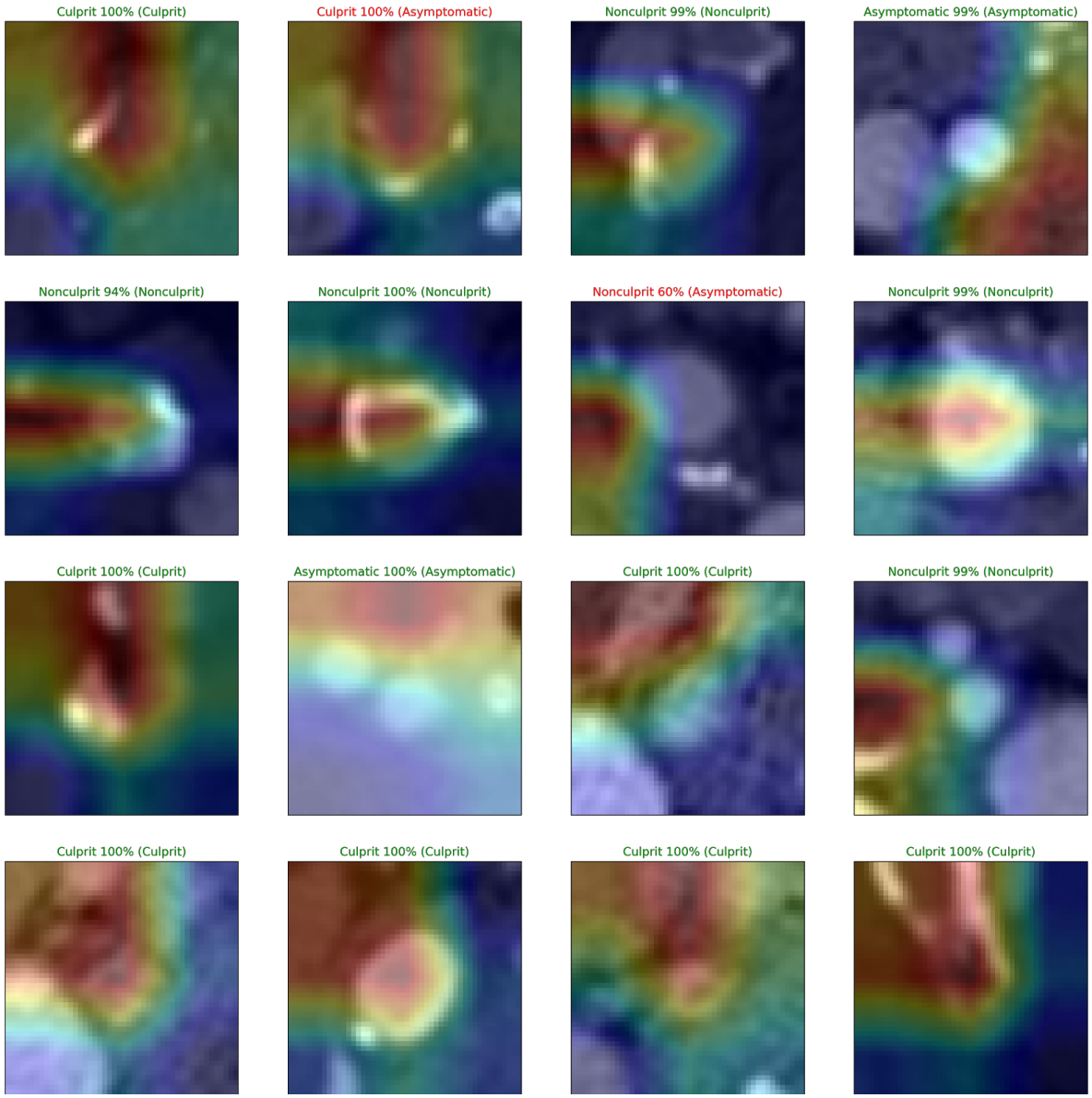


Figure 6.98 Multi-class Internal Validation Data Class Activation Maps

The internal validation dataset consisted of 420 carotid image patches. The model's predictions are the first label and the ground truth labels are in brackets. Green indicates the correct predictions and red indicates the incorrect predictions. The probability or 'confidence' of the model associated with the classification decision per image, is provided as a percentage. Areas of maximal activation are highlighted in red and regions of minimal activation are highlighted in blue within the class activation heatmap.

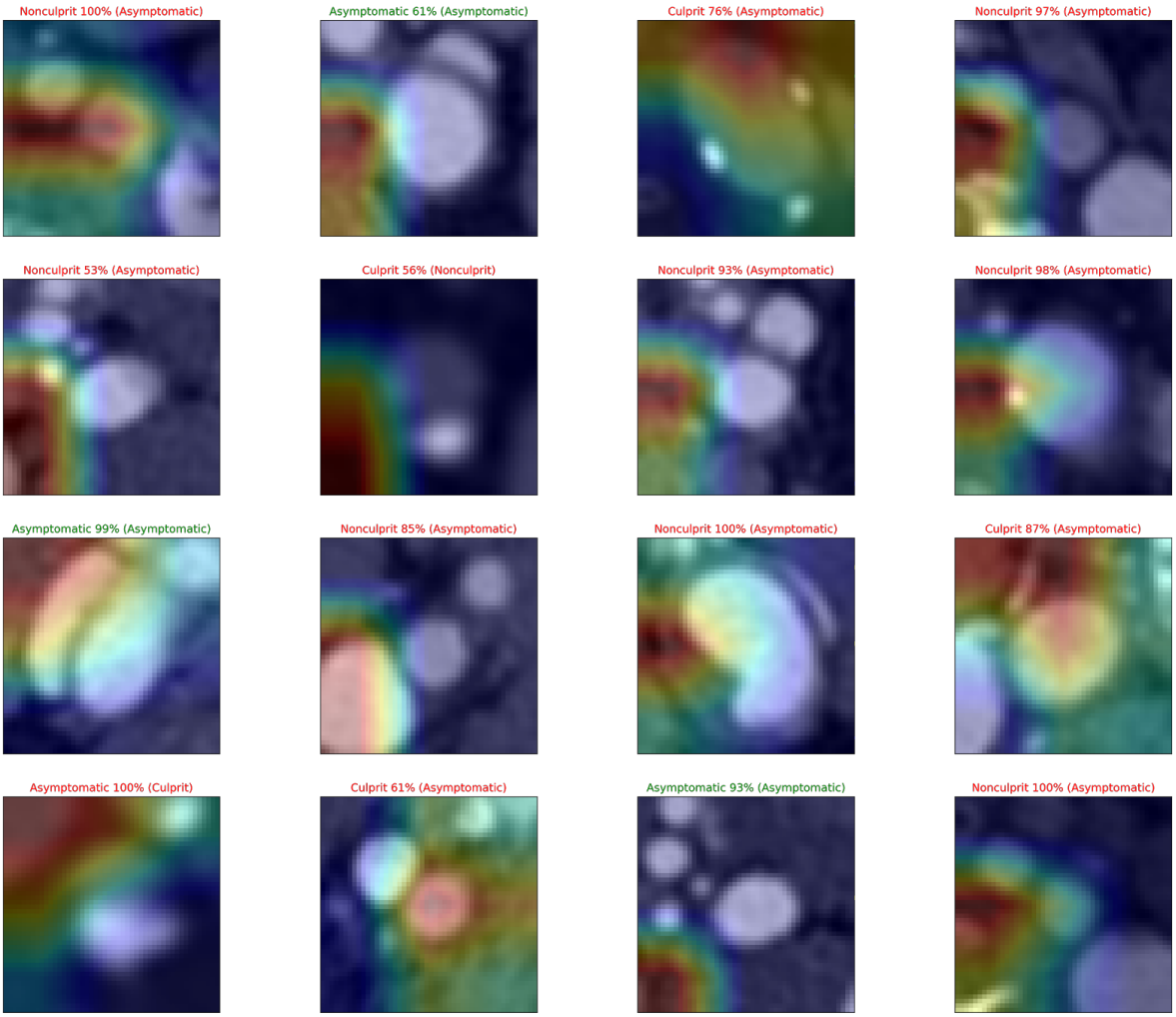
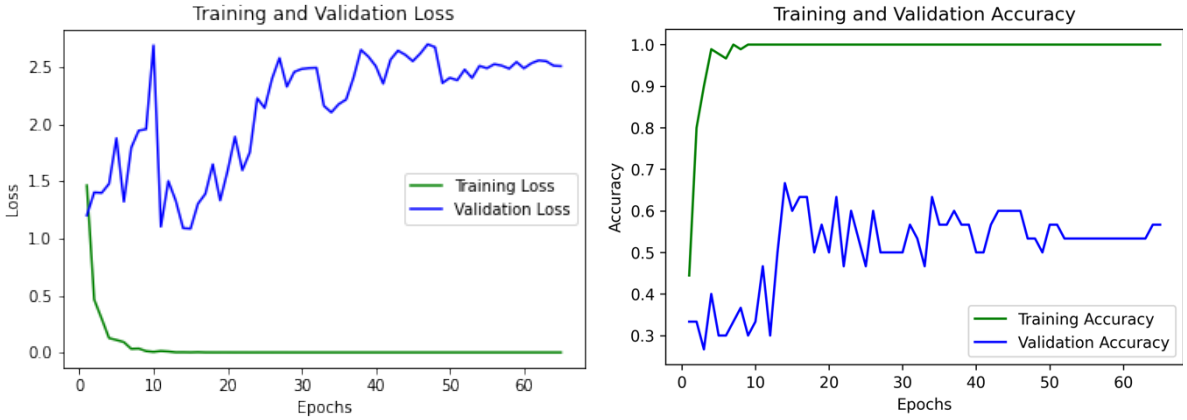


Figure 6.99 Multi-class External Validation Data Class Activation Maps

The external validation dataset consisted of 168 carotid image patches derived from 12 carotid arteries. This figure provides overlays of Grad-CAM class activation heatmaps on the carotid image patches of external validation data. Image label format: (1) model's prediction of carotid type, (2) probability for that class prediction and (3) ground truth label of carotid type. Those in green were correct predictions, those in red were incorrect.

6.3.2 Multi-class Classification with a 3D Convolutional Neural Network

A 3D CNN model (see Figure 6.88 for the model architecture), was trained for multi-classification of carotid artery types. The dataset consisted of 40 asymptomatic, 40 culprit and 40 non-culprit carotid 3D volumes (each formed by the fusion of 14 carotid image patches, described in section 6.2.4.3). The data was split such that 75% was used for training (90 carotid 3D volumes) and 25% for internal validation (30 carotid 3D volumes). In addition, there were 12 carotid 3D volumes (that correspond to the 12 carotid arteries used for external validation in section 6.3.1.3) that was used for external validation in this task. Figure 6.100 provides the training and internal validation set loss and accuracy curves.



A) Loss curves **B) Accuracy curves**

Figure 6.100 Loss and Accuracy Curves for Training and Validation Sets in Multi-class Classification with a 3D CNN

The 3D CNN was quick to reach 100% accuracy on the training set, but not on the internal validation set i.e. it quickly overfit on the training dataset. Evaluating the model on the (1) training dataset resulted in a mean accuracy of 100%, on the (2) internal validation set resulted in a mean accuracy of 67% and (3) on the external validation set resulted in a mean accuracy of 42% and an AUC of 0.64. The predictions, ground truth labels and the 3D CNN's confidence in its respective classification decisions for 3D carotid volumes from the training set, internal validation set and the external validation set are provided in Figure 6.101, Figure 6.102, Figure 6.103 respectively.

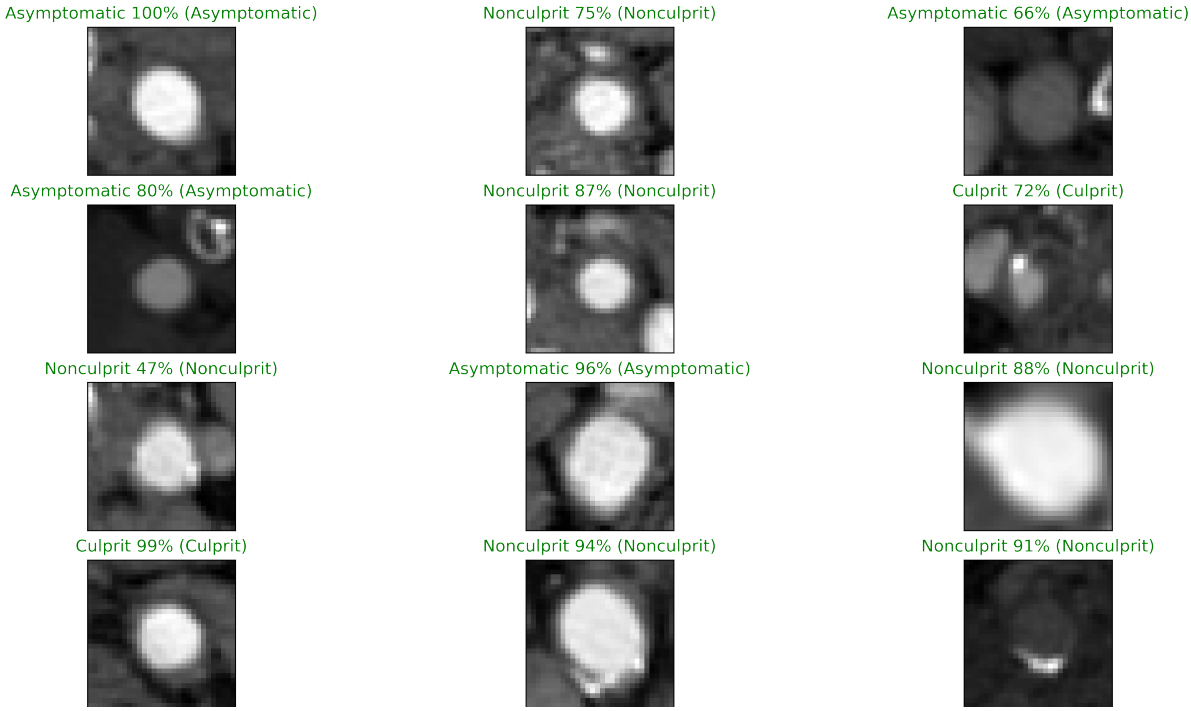


Figure 6.101 3D CNN Predictions on the Training Dataset

The training dataset consisted of 90 carotid 3D volumes. Image label format: (1) model's prediction of carotid type, (2) probability for that class prediction as a percentage and (3) ground truth label of carotid type. The image is one slice from the carotid 3D volume.

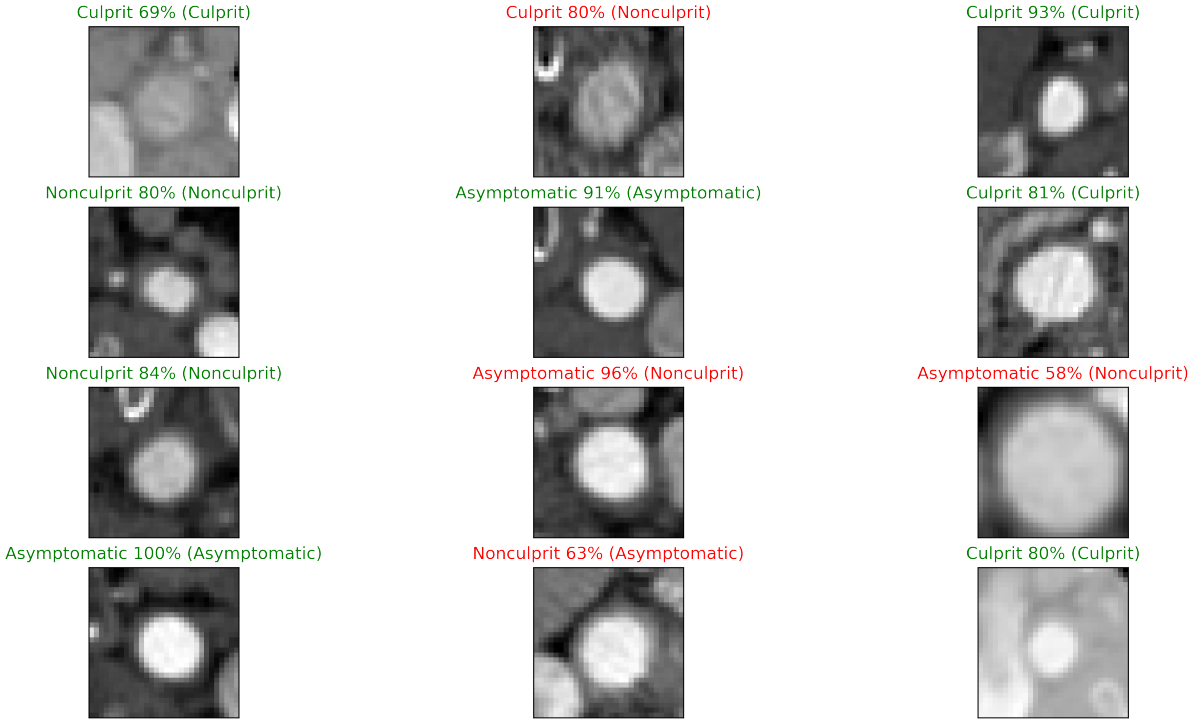


Figure 6.102 3D CNN Predictions on Validation Data Subset

The internal validation set consisted of 30 carotid 3D volumes. Green indicates correct predictions and red indicates the incorrect predictions.

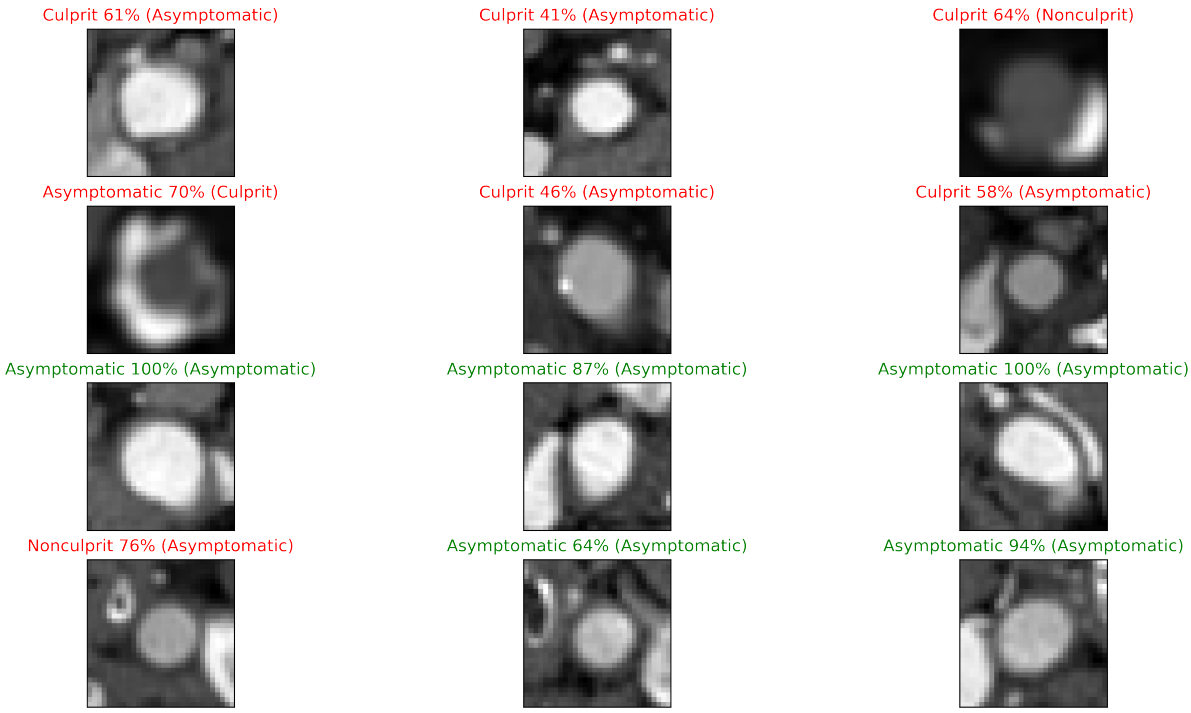


Figure 6.103 3D CNN Predictions on External Validation Data

The external validation dataset consisted of 12 carotid 3D volumes. Image label format: (1) model's prediction of carotid type, (2) probability for that class prediction as a percentage and (3) ground truth label of carotid type.

The confusion matrix for the model's performance on the internal validation set is provided in Table 6.104 and for the external validation set in Table 6.105.

Table 6.104 Multi-class Classification with a 3D CNN: Internal Validation Confusion Matrix

Confusion matrix	**Predicted Values**		
True Values	**Asx**	**CC**	**NC**
Asx	**6**	1	3
CC	0	**10**	0
NC	2	4	**4**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the CNN classifier are highlighted in red.

Table 6.105 Multi-class Classification with a 3D CNN: External Validation Confusion Matrix

Confusion matrix	Predicted Values		
True Values	**Asx**	**CC**	**NC**
Asx	**5**	4	1
CC	1	**0**	0
NC	0	1	**0**

Asx, asymptomatic; CC, culprit; NC, non-culprit carotid arteries. The incorrect predictions of the CNN classifier are highlighted in red.

The sensitivity and specificity respectively for each carotid artery class was as follows: Asx (50%, 50%), CC (0%, 58%) and NC (0%, 91%).

6.3.3 Stress Testing with 2D and 3D CNNs

In this multi-class classification task, we experimented with applying various image operations to the 2D carotid image patches and the 3D carotid volumes in order to assess the impact on the deep learning models' classification decisions. This was a preliminary investigation into the robustness of the deep learning model.

6.3.3.1 Stress Testing with 2D CNNs

Figure 6.104 demonstrates that the 2D CNN (modified VGG16 network) trained using transfer learning for multi-class classification was relatively robust to input image perturbations. The greatest impact on the probabilities for the respective carotid artery classes resulted from horizontal flipping.

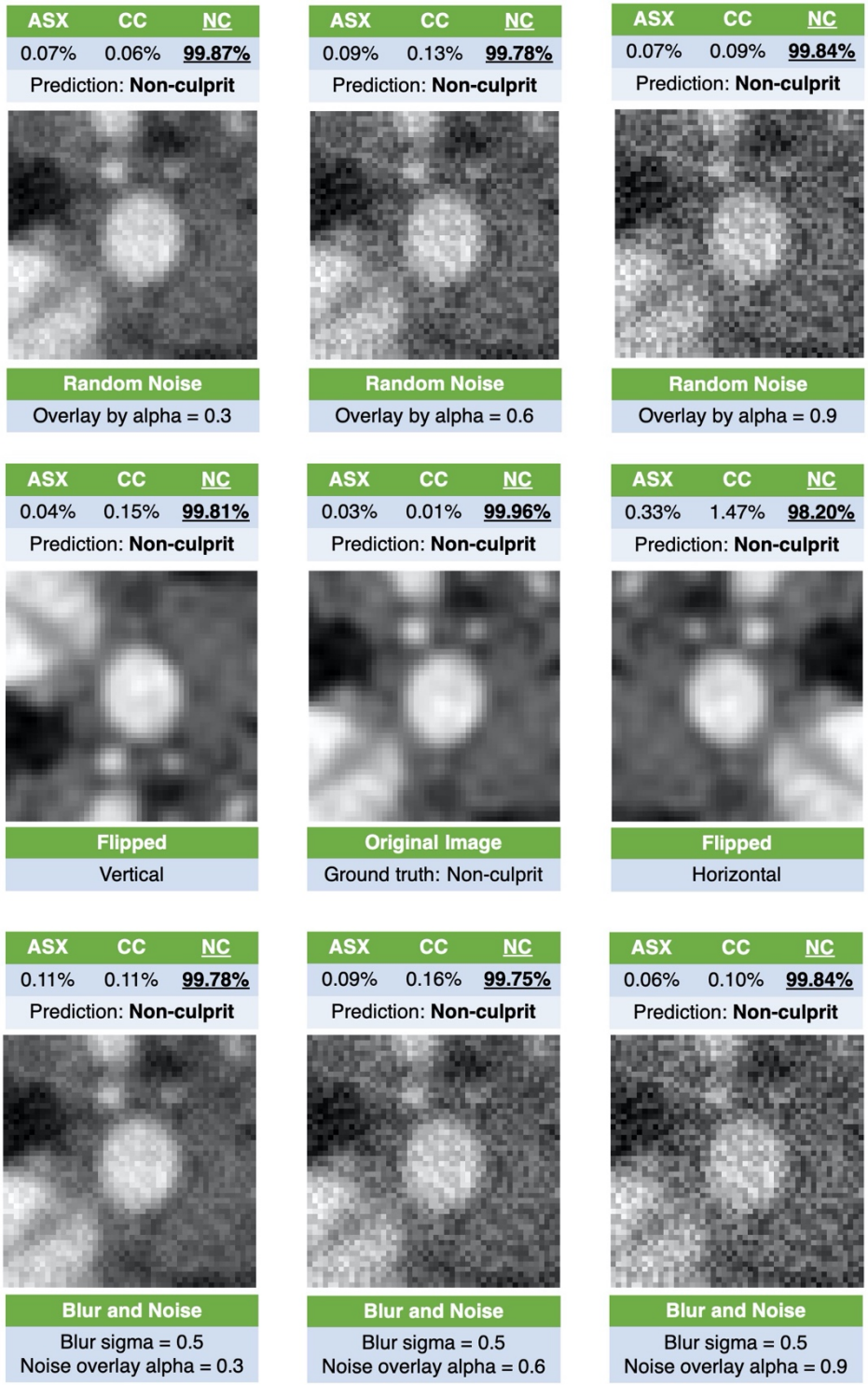


Figure 6.2 Stress testing the 2D CNN with image manipulations: Flipping, Noise and Blur

The original input image (carotid 2D image patch) is in the centre of the figure. The ground truth label is non-culprit carotid artery. The deep learning model's classification prediction and associated probability per carotid artery type are at the top of each image. ASX, asymptomatic; CC, culprit; NC, non-culprit carotid artery. Random noise is sampled from the uniform distribution [0,1]. Blur refers to Gaussian Blur.

6.3.3.2 Stress Testing with 3D CNNs

Figure 6.105 demonstrates that increasing the amount of Gaussian blur within the input image, led to a decrease in the probability that the model classified the input as 'non-culprit' and using a Gaussian blur filter with sigma 1.0 was sufficient to change the classification decision from 'non-culprit' to 'culprit' carotid artery. In addition, it demonstrated that horizontally flipping the image could also change the classification decision from 'non-culprit' to 'culprit' carotid artery.

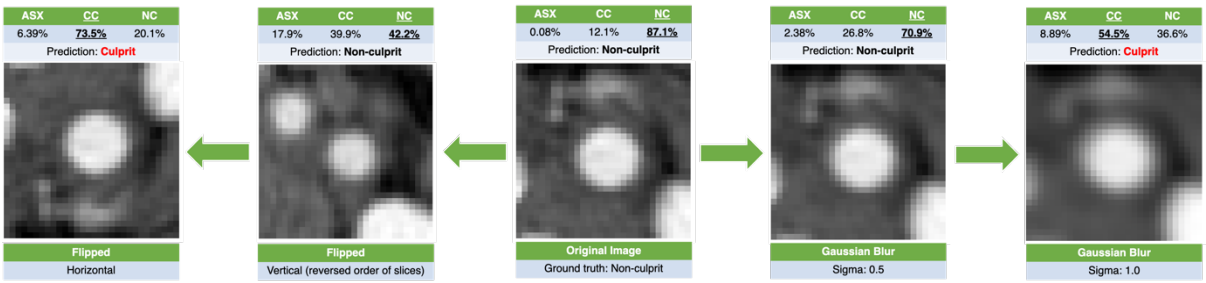


Figure 6.105 Stress testing the 3D CNN with image manipulations: Flipping and Blur

The original input image (carotid 3D volume, although we only visualise one slice from the volume in this figure) is in the centre of the panel. The deep learning model's classification prediction and associated probability per carotid artery type are at the top of each image. ASX, asymptomatic carotid artery; CC, culprit carotid artery; NC, non-culprit carotid artery. From left to right, the following image manipulations were applied to the original centre image: (1) horizontal flip, (2) vertical flip, (3) original image (ground truth label is non-culprit), (4) Gaussian blur (sigma = 0.5) and (5) Gaussian blur (sigma = 1.0). Red indicates a change in the model's classification decision.

Figure 6.106 demonstrates that a combination of noise and blurring could change the classification decision from 'non-culprit' to 'asymptomatic' carotid artery.

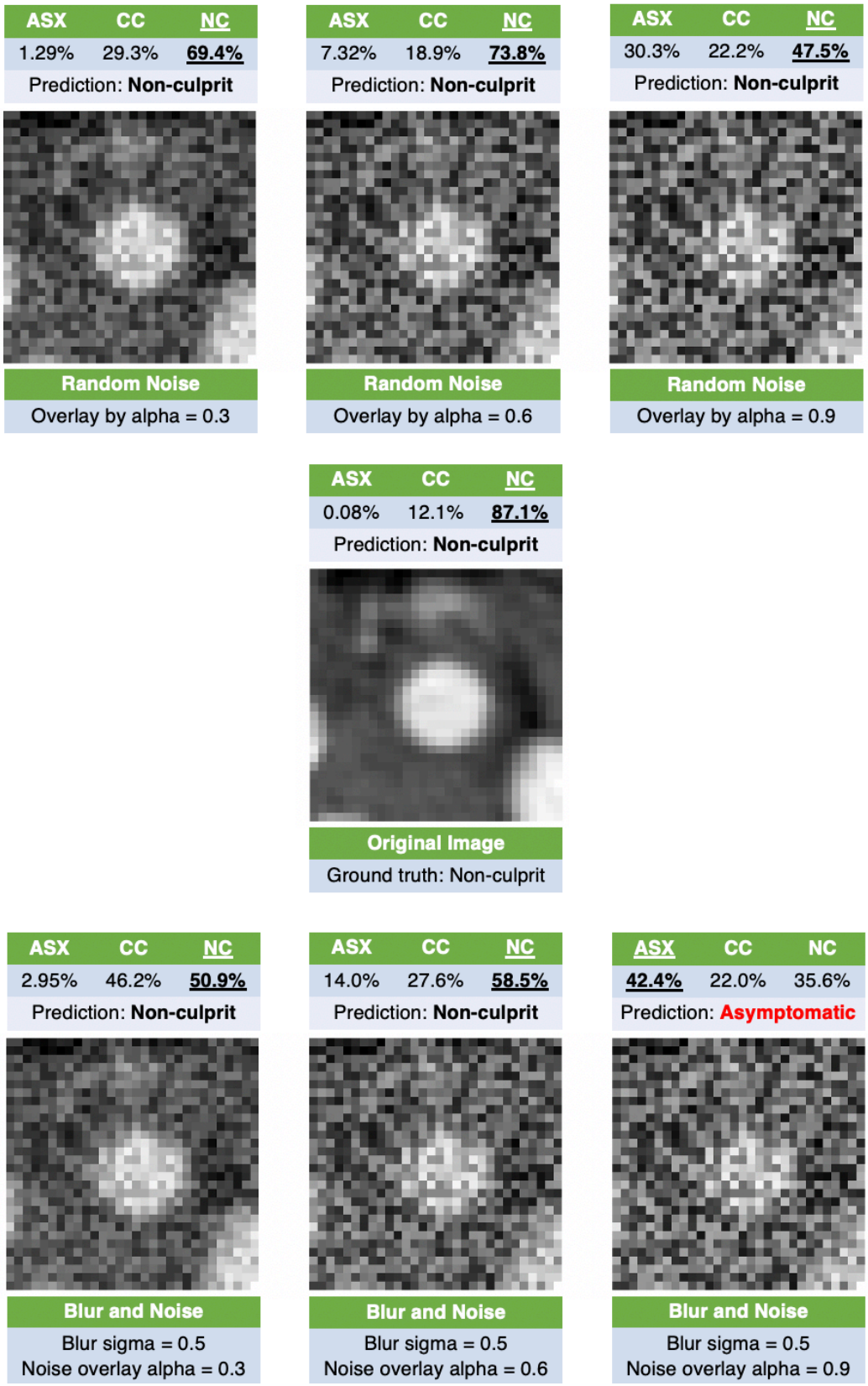


Figure 6.106 Stress testing the 3D CNN with image manipulations: Noise and Blur

The original image is at the centre of the figure (ground truth label is non-culprit carotid artery). Random noise is sampled from the uniform distribution [0,1]. Blur refers to Gaussian Blur. The model's predictions are at the top of the image along with the associated probabilities per carotid artery class. Red indicates a change in classification decision.

6.4 Discussion

Convolutional neural networks are widely employed in computer vision and deep learning applications within the medical field such as in mammography interpretation[114]. We investigated the ability of deep learning to correctly classify CT angiography image patches of the carotid arteries, without the need for prior feature engineering and feature selection.

6.4.1 Binary classification (Transfer Learning)

In the binary classification of culprit versus non-culprit carotid arteries using the carotid image patch from the carotid bifurcation only, we found that using transfer learning i.e. VGG16 model initialised with pretrained ImageNet weights resulted in greater predictive performance than using a deep learning model that had to be trained from scratch. This suggests that even though the original ImageNet dataset was comprised of coloured (RGB) images representing objects in the real world, the features the model learnt from that vast dataset had some transferable utility in classifying carotid arteries in medical images. This was further supported by the visualisations of the outputs (i.e. activations) from different layers within the convolutional neural network. These visualisations provided an insight into the visual patterns that each filter in the CNN was receptive to. In the earlier layers of the CNN, we could see that the VGG16 filters could extract the contour of the carotid artery wall and areas of macrocalcification. Therefore, certain filters learnt from ImageNet are applicable to carotid imaging. However, not all filters were relevant to carotid imaging, particularly those in the later layers of the CNN which represented more abstract concepts and did not activate with the carotid image patch inputs.

For the binary classification of whether a carotid artery was derived from an asymptomatic versus a symptomatic patient, a higher internal validation accuracy was achieved compared with the discriminating between culprit versus non-culprit carotid arteries. This reflects the results from previous chapters using radiomic features in which more differences exist between asymptomatic and symptomatic carotid arteries than between culprit and non-culprit carotid arteries. For both of these binary classification tasks, the input image was a single carotid image patch and the predictive performance achieved was greater than chance in each case. This suggests that there is a detectable signal even within a single slice of the carotid artery that can inform us about the clinical status of the entire carotid artery.

6.4.2 Multi-class Classification (Transfer Learning)

In this multi-class classification task, the data used for training the deep learning model and for external validation followed the same set up as in Chapter 5 which evaluated the use of a feature-based machine learning approach with carotid CTA radiomic features. Whilst the distribution of carotid artery types was different between the training dataset and the external dataset (the training data had an equal distribution of classes, whereas the external dataset was skewed towards the asymptomatic class), the radiomics feature-based approach performed well on the external dataset. However, this was not the case for a deep learning approach, whether transfer learning was used (section 6.3.1.3) or not (section 6.3.2). This likely relates to the higher model complexities of deep learning approaches compared with feature-based machine learning and consequent overfitting to the training data.

When using transfer learning in the multi-class classification task, the performance on the training and internal validation datasets were better than chance. However, the performance on the external validation dataset was worse than chance. Visualisations of the class activation maps for the training dataset, validation set and the external validation dataset revealed that in general when the model had managed to correctly localise the carotid artery, there was better classification performance and this was more likely to be the case for the training data carotid arteries. The class activation maps for the external validation dataset revealed that the model was not activating around the carotid artery and so the model had not properly learnt the underlying patterns relevant to carotid artery classification.

6.4.3 Multi-class Classification with 3D CNN

In the 3D CNN approach, this enabled the model to receive input that contained information from multiple slices of the carotid artery, rather than a single slice as in the approaches discussed above. However, this reduced the overall sample size of the dataset and required a custom 3D CNN model to be built and randomly initialised rather than using transfer learning, since 3D CNN architectures with pretrained weights are not currently available in Keras.

From the training and validation loss and accuracy curves, we can see that the simple 3D CNN model was quick to overfit on the training data without much success on the internal validation

set. This demonstrated that the 3D CNN was sufficiently complex to learn from the data, but also that perhaps it was too complex, or the sample size was too small.

6.4.4 Stress Testing

For the multi-class classification tasks, the feasibility of robustness analysis as applied to deep learning was investigated by applying different image manipulations to an input image that the model had initially classified correctly. This stress testing was inspired by the ideas introduced in Chapter 4 which investigated the robustness of the radiomic features. However, as using a deep learning approach takes longer and is more intensive computationally than a radiomics feature-based machine learning approach, the stress testing in this chapter was not a systematic investigation but rather a preliminary feasibility exercise. The results demonstrated that subtle changes to the input image (particularly in the 3D CNN case) was sufficient to induce misclassification, for example by increasing the level of image blurring or orientation of the image. This highlights the potential vulnerabilities of a deep learning approach that may not be obvious with a simple train-test split evaluation of a deep learning model. Furthermore, this is in line with the findings of Szegedy and colleagues who found that certain imperceptible perturbations (termed adversarial attacks) of the input image could change the deep learning model's classification decisions[254].

6.4.5 Limitations and Future Work

In terms of deep learning applications whereby the ImageNet dataset consisted of 1.2 million images, the carotid image dataset is comparatively small. This small sample size increased the risk of the model overfitting on the training data and was demonstrated when the training loss curves continued to decrease as the validation loss curves diverged and increased. To try and overcome the limitation of the small dataset, we tried to leverage the power of transfer learning. However, there are inevitably differences between the everyday RGB images found in ImageNet and medical images which limited the success of the deep learning approach in this chapter.

The underlying assumption for deep learning is that there are detectable, perhaps subtle changes in imaging that are predictive of future events or are relevant to the clinical problem. The challenge is for the deep learning model to learn the correct discriminative regions from

the image inputs. In this study, even though the input to the deep learning models had been restricted to the carotid image patch, there remained challenges for the deep learning algorithm as there could have been the presence of the jugular vein, bone and muscle for example. The poor performance of the models on the external validation dataset suggest that we need to further confine the search space, and this could be achieved by introducing a carotid artery segmentation step into the deep learning workflow. For example, the AI framework utilised by De Fauw and colleagues in 2018 in order to classify OCT images in terms of ophthalmological diagnoses involved firstly a segmentation network followed by a classification network[255]. Alternatively, a multi-task learning approach could be explored which involves training the same model to do different tasks at the same time. The underlying idea is that there will be information overlap between the tasks that can help improve the model's ability at each individual task i.e. training an image-classification model jointly with an image-segmentation model that shares the same convolutional base would result in a model that is better at both tasks, compared to a model that was trained only on one specific task alone[256].

With regards to the external validation dataset used in this chapter, the limitations of the different class distribution between the training set and the external validation set has previously been discussed in Chapter 5. This approach enabled a direct comparison between the radiomics feature-based machine learning approach in multi-class classification in Chapter 5 with the deep learning approach in this chapter. The performance of the deep learning approach indicates a need for more data collection and future work could explore a federated machine learning approach. Federated machine learning is also known as collaborative learning and reverses the paradigm of bringing the data to the model into bringing the model to the different local data sources. This approach enables the model to be trained and evaluated without the need for data exchange between centres and research is underway into making this approach scalable and more efficient[257].

The performance of the deep learning model could further be improved with hyperparameter tuning, which is both an art and a science. In addition, the field of deep learning is rapidly developing, with constantly improving versions of activation functions and optimisers that could be incorporated. For example, in this study we used RMSprop as the optimiser, but there exist improvements to this optimiser such as the Adam optimiser[258] which adds in the concept of momentum[259]. In the future, by following these steps, we will hopefully be able to train a model that has learnt suitable and relevant discriminative regions within the carotid images

with the ability to generalise on unseen data. We can then develop the stress testing introduced in this chapter into a systematic robustness analyses to gain greater insight and confidence in the deep learning model's classification decisions.

6.5 Conclusion

Neural networks learn a mapping from the input data (training data) and the output variable that we want to predict (e.g. the carotid artery type). They can learn different representations from the training data and discover different methods to map the input from the output. Mathematically, they are capable of learning any mapping function (universal approximation theorem)[141]. Whilst these properties provide great opportunities for computer vision classification tasks, they also present significant challenges. These include identifying a suitable model architecture and complexity for the classification task, hyperparameter tuning, use of suitable representative datasets and overcoming the issue of overfitting and generalisability. Whilst the 2D deep learning models in this chapter performed well in the training and internal validation datasets, they did not in the external validation test set. A preliminary stress testing exercise of the deep learning model demonstrated vulnerabilities to the amount of image noise or how blurry the image was in determining the carotid artery class. Overall, more data and refinement of the deep learning methods is required before a deep learning approach can outcompete a radiomics feature-based machine learning approach in classifying carotid artery status from carotid CTA images.

Chapter 7: General Discussion, Future Directions and Conclusion

Chapter summary:

This chapter provides a summary of the main findings of this research and explores its implications for vascular CTA imaging. In addition, an outline of future work and possible future directions are discussed.

7.1 Summary and implications of the main findings

The work in this thesis aimed to evaluate the feasibility of applying radiomics and machine learning to carotid CT angiography scans for the identification of culprit carotid arteries in the context of stroke or TIA, and to investigate the biological associations of radiomic features with atherosclerotic pathophysiology.

The work described in chapters 2 and 3 demonstrated that carotid radiomic features are quantitative features that can be derived from carotid CT angiography scans and unenhanced CT images. Radiomic features derived from carotid CTA scans had better discriminatory ability compared with those from unenhanced CT scans for the identification of different carotid artery types, namely (1) culprit, (2) non-culprit and (3) asymptomatic carotid arteries. The results indicated clear differences in the radiomic profiles of culprit versus non-culprit carotid arteries, as well as differences between asymptomatic carotids and symptomatic carotid arteries. These differences were not limited to first order texture features but extended to higher-order radiomic features derived from matrices that considered the spatial interrelationships between pixels. The differences between asymptomatic versus symptomatic carotid arteries were greater than the differences between culprit versus non-culprit carotid arteries. Using a multi-slice approach, as opposed to a single-slice approach, with resegmentation (i.e. a restricted algorithm that excludes carotid calcium) was the most effective approach for revealing differences between the 3 carotid artery types. In particular, this was most beneficial in revealing differences between carotid arteries with mixed plaque types and in those with less than 50% carotid stenosis.

Chapter 4 illustrated a systematic approach to assessing the robustness of radiomic features to ROI perturbations. These perturbations were used to mimic the ROI placement over- and

under-estimations introduced by human subjectivity in clinical practice when using manual delineation methods. The results showed that not all of the extracted radiomic features were robust against these perturbations and that the proportion of highly robust radiomic features was dependent on the image pre-processing methods and image quantisation methods used. Multi-slice analysis produced a higher proportion of radiomic features with excellent robustness than single-slice analysis and radiomic features identified as highly robust in one image setting was not necessarily robust in another setting. A set of robust and non-redundant radiomic features for the differentiation of culprit and non-culprit carotid arteries was identified for multi-slice and single-slice analysis, with and without resegmentation. These radiomic feature sets demonstrated superior predictive performance to carotid calcification alone as a predictor in this binary classification task.

Chapter 5 extended the binary classification task into a multi-class task to discriminate between the 3 carotid artery types. The predictive ability of carotid calcification, PET information using several tracers and radiomic features both alone and together were investigated. The results demonstrated that radiomic features had better predictive performance than carotid calcification and PET information in feature-based machine learning. Importantly, these carotid CTA radiomic features were related to markers of inflammation, defined by relationships with PET and immunohistochemistry. Radiomic features associated with low attenuation values were positively correlated with FDG, such as GLDM: Low Grey Level Emphasis. Measures of local homogeneity within an ROI such as GLCM: Inverse Difference Normalised were correlated with DOTATATE. Radiomic features of non-uniformity were strongly correlated with FMISO, such as NGTDM: Complexity and NGTDM: Strength, whilst radiomic features with resegmentation were associated with %CD68 such as GLDM: Dependence Entropy and First Order: Range. This suggests that they do reflect relevant features of atherosclerosis and demonstrates the biological plausibility of using carotid radiomic features for classification and to provide insights into the disease process.

The work in chapter 6 went beyond the data mining capabilities of radiomics and explored a deep learning approach for the extraction of carotid imaging features not limited to those defined in either TexRAD or PyRadiomics. The results demonstrated that although deep learning was a viable method for learning new features from carotid imaging data, the risk of overfitting was greater than that with feature-based machine learning. This was likely because the model's complexity and flexibility was much greater. Subjecting the input images to

varying amounts of noise and blurring revealed vulnerabilities in the deep learning model; these changes could change the model's carotid artery type classification decision. Even with the use of transfer learning to overcome the limited dataset available, our results indicated that 1) more data and 2) refinement of the deep learning methods would be needed for the deep learning approach to outperform feature-based radiomics for classifying carotid artery status from carotid CTA images.

Together, the findings of this work imply that additional information can be mined from carotid CT angiograms by using radiomics, going beyond luminal stenosis. Since CT angiography forms part of the clinical workflow in ischaemic cerebrovascular management, the application of radiomics to the carotid arteries would not require any further imaging or radiation to the patient. Whilst carotid CTAs are known to provide anatomical information about the carotid arteries, the associations between carotid CTA radiomic features and markers of carotid inflammation suggest that radiomics enables potentially useful functional information about atherosclerotic activity to be extracted from the CTA scans. With further refinement, CTA radiomics could compete with or even outcompete vascular PET imaging to provide an indication of atherosclerotic activity and could be useful in re-classifying subgroups of patients. For example, those with increased disease activity as indicated by their carotid radiomic profiles might benefit from more aggressive medical therapies than those with less disease activity.

According to Prescott et al., there are two necessary criteria for the validation of a quantitative imaging biomarker: (1) 'the presence of the quantitative imaging biomarker is closely coupled or linked to the presence of the target disease or condition' and (2) 'the detection and quantitative measurement of the quantitative imaging biomarker are accurate, reproducible and feasible over time'[66]. Overall, the findings in this thesis have begun to address those two criteria and suggest that carotid radiomics provides a non-invasive strategy to study carotid artery inflammation with the potential to act as an imaging biomarker in identifying plaque vulnerability. This project represents the initial proof-of-concept stage which forms the basis for future work before any clinical deployment.

7.2 Future work and directions

The application of radiomics to vascular imaging has begun but is far less developed than other areas such as oncology[189]. The radiomics quality score developed by Lambin et al. [260] provides a framework for the evaluation of radiomic studies in their readiness for clinical translation. The scoring system consists of 16 different criteria including the provision of well-documented image protocols, imaging at multiple time points and investigating biological correlates to the radiomic features. Whilst the work in this thesis has demonstrated both feasibility and a sound underpinning in biology for radiomics, it scores 16 out of a possible 36 points according to the radiomics quality score, indicating that further work is needed in several areas.

One of the most time-consuming aspects of the current workflow is the manual segmentation of the carotid arteries to produce the regions-of-interest from which radiomic features are extracted. This was also a source of intra-rater and inter-rater variability. Whilst we have demonstrated that there are radiomic features that are robust to inter-rater variability in ROI delineation, not all radiomic features were immune to it. In oncological radiomic studies, semi-automated and automated segmentation approaches have been investigated[261]. For example, Gu et al. found that using a semi-automated method for the delineation of lung tumours from CT scans provided a way for different readers to be able to obtain nearly identical segmentation results using a single click. By doing so, they reported better results than by using manual segmentation[262]. Therefore, exploration of semi-automated and automated methods for carotid artery segmentation should reduce the time taken for image analysis and provide reproducible, consistent segmentations. Taking this further, in order to enhance the usability and efficiency of image analysis for the end-user (clinicians), the development of an integrated and automated software solution that combines all the steps of the workflow into one represents the ultimate aim. This would ideally handle carotid segmentation, radiomic feature extraction and machine learning analysis all in one and would enhance the scalability of this workflow.

Additionally, the radiomic features extracted in this study were not limited simply to the carotid plaque but incorporated information about the entire artery. This was because segmenting the carotid artery was more straightforward and reproducible. However, it would be interesting to investigate the predictive value of radiomic features derived from the carotid plaque alone, as well as those derived from the surrounding perivascular fat of carotid arteries. For example, in coronary arteries, phenotypic changes in the perivascular fat outside the artery are associated

with atherosclerotic inflammation as a result of pro-inflammatory cytokine release[219]. Radiomic analysis of the pericoronary adipose tissue has recently been shown to be able to discriminate patients with acute myocardial infarction from asymptomatic patients[220], and improve risk prediction for adverse clinical events in patients with low coronary calcium scores[263].

The radiomic analyses conducted in this study originated from a single institution and a single scanner. Further research is needed to examine whether these radiomic features are scanner or centre-specific or whether these features have predictive value that is generalisable to multi-centre data and are scanner agnostic. This is a recognised phenomenon in machine learning whereby scanner-specific bias may lead to overly optimistic performance estimates, for example, Glocker et al. found that using current pre-processing methods for multi-centre data harmonisation, machine learning classifiers were still able to easily determine the origin of multi-site neuroimaging data[264]. In order to investigate this further, future work should aim to acquire multi-centre data from prospective clinical outcome studies. These longitudinal studies would also enable time series information to be incorporated into the predictive models such as time-to-event data for better personalised risk prediction.

In oncology, repeated radiomic analysis has been investigated to monitor pharmacologic responses to therapy[265]. Similarly, FDG PET has been investigated for monitoring the therapeutic effect of anti-inflammatory drugs in atherosclerosis[266]. Since the findings in this thesis suggest that carotid radiomic features have better predictive value in identifying culprit carotid arteries compared with FDG PET and that radiomic features have associations with atherosclerotic processes, future studies should consider the use of radiomic features for pharmacologic response monitoring. Additionally, this would provide test-retest data to assess the reproducibility and accuracy of carotid radiomics features over time.

Future work might also explore radiomic features extracted after additional filtration and transformation steps, such as fractal analysis, which reveals information about the shape and complexity of the analysed image[261] or wavelet decomposition which extracts features from different frequency bands[267] and which has been employed in oncology to fuse texture characteristics from two imaging modalities e.g. CT and PET[268]. In addition, radiomics has scope for integration with other big data sources. In oncology, radiogenomics[61] is advancing the field into an era of precision medicine[269]. Radiomics provides imaging information that is

spatially and temporally specific to the index tumour that can supplement that gained from genomic analysis, such as mutation load[270]. This approach may be transferable to the vasculature, particularly using data from biobanks, such as the UK Biobank, that contains imaging data and genome-wide genotyping data for 500,000 participants and subsequent clinical outcomes[271].

Within the scope of stroke medicine, future work could explore the ability of carotid radiomic features to predict brain MRI features and micro-vessel disease severity for risk stratification, since there seems to be an association between carotid artery plaque instability (as indicated by higher ultrasound carotid artery strain indices) and white matter ischaemic injury (as determined by T2-weighted brain MRI imaging)[272]. In addition, current projects within the Department of Cardiovascular Medicine are acquiring larger datasets of carotid samples from carotid endarterectomy. From these samples, high resolution images of the histological slices could be obtained for textural analysis and the histological radiomic features could be correlated with those obtained from carotid CTA imaging.

Finally, radiomics can extend to other vascular beds. For example, the femoral arteries are of similar calibre to carotid arteries and therefore the algorithms investigated in this thesis would be applicable to the investigation of peripheral vascular disease. Future work could also extend to the coronary arteries and abdominal aortic aneurysms. Radiomic analyses of these vascular beds would allow us to study the similarities and differences between the radiomic feature values of the respective vasculature and give us a better idea about whether CTA radiomics are reflecting local inflammation specific to the carotid arteries or general inflammation within the patient since atherosclerosis is a systemic process. Ultimately, prospective multicentre studies would be required to validate these findings, ideally with a workflow that comprises embedded radiomics using automatic vessel and plaque segmentation.

7.3 Conclusions

The central aim of this thesis was to evaluate carotid atherosclerosis using radiomics and machine learning analysis of carotid CT angiography scans from symptomatic and asymptomatic patients in the context of stroke and TIA. This research demonstrates the feasibility of using carotid CTA radiomics for the identification of different carotid artery types: culprit, non-culprit and asymptomatic carotid arteries. The radiomic features were shown to be associated with atherosclerotic processes as determined by PET imaging and immunohistochemistry, and radiomic features sets were shown to be robust against perturbations in the radiomics workflow. This highlights the potential of carotid CTA radiomic features as useful quantitative imaging biomarkers for better stroke risk stratification and the potential personalised tailoring of patient management. Future work will explore methods to optimise the carotid CTA radiomics workflow and the acquisition of prospective data for clinical validation in order to support the transition of radiomics into the clinical practice of cardiology and stroke medicine.

Bibliography

1. Nowbar, A. N., Gitto, M., Howard, J. P., Francis, D. P. & Al-Lamee, R. Mortality From Ischemic Heart Disease: Analysis of Data From the World Health Organization and Coronary Artery Disease Risk Factors From NCD Risk Factor Collaboration. *Circ. Cardiovasc. Qual. Outcomes* **12**, (2019).
2. Bhatnagar, P., Wickramasinghe, K., Wilkins, E. & Townsend, N. Trends in the epidemiology of cardiovascular disease in the UK. *Heart* **102**, 1945–1952 (2016).
3. Roth, G. A. *et al.* Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. *J. Am. Coll. Cardiol.* **70**, 1–25 (2017).
4. Health Matters: Preventing cardiovascular disease - Public health matters. https://publichealthmatters.blog.gov.uk/2019/02/14/health-matters-preventing-cardiovascular-disease/.
5. Patel, A. *et al.* Estimated societal costs of stroke in the UK based on a discrete event simulation. *Age Ageing* **49**, 270–276 (2020).
6. Kendir, C., van den Akker, M., Vos, R. & Metsemakers, J. Cardiovascular disease patients have increased risk for comorbidity: A cross-sectional study in the Netherlands. *Eur. J. Gen. Pract.* **24**, 45–50 (2018).
7. Li, M. *et al.* Cardiovascular disease potentially contributes to the progression and poor prognosis of COVID-19. *Nutr. Metab. Cardiovasc. Dis.* **30**, 1061–1067 (2020).
8. Clerkin, K. J. *et al.* COVID-19 and Cardiovascular Disease. *Circulation* **141**, 1648–1655 (2020).
9. Bentzon, J. F., Otsuka, F., Virmani, R. & Falk, E. Mechanisms of Plaque Formation and Rupture. *Circ. Res.* **114**, 1852–1866 (2014).
10. Tarkin, J. M. *et al.* Imaging Atherosclerosis. *Circ. Res.* **118**, 750–769 (2016).

11. Libby, P. Superficial erosion and the precision management of acute coronary syndromes: not one-size-fits-all. *Eur. Heart J.* ehw599 (2017) doi:10.1093/eurheartj/ehw599.
12. Sandfort, V., Lima, J. A. C. & Bluemke, D. A. Noninvasive Imaging of Atherosclerotic Plaque Progression: Status of Coronary Computed Tomography Angiography. *Circ. Cardiovasc. Imaging* **8**, (2015).
13. Bom, M. J. *et al.* Early Detection and Treatment of the Vulnerable Coronary Plaque: Can We Prevent Acute Coronary Syndromes? *Circ. Cardiovasc. Imaging* **10**, (2017).
14. Arbab-Zadeh, A. & Fuster, V. The Myth of the "Vulnerable Plaque". *J. Am. Coll. Cardiol.* **65**, 846–855 (2015).
15. Skeoch, S. *et al.* Imaging atherosclerosis in rheumatoid arthritis: evidence for increased prevalence, altered phenotype and a link between systemic and localised plaque inflammation. *Sci. Rep.* **7**, 827 (2017).
16. Feigin, V. L. *et al.* Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *The Lancet* **383**, 245–255 (2014).
17. State of the Nation: stroke statistics. *Stroke Association* https://www.stroke.org.uk/resources/state-nation-stroke-statistics (2015).
18. Sacco, R. L. *et al.* An Updated Definition of Stroke for the 21st Century: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* **44**, 2064–2089 (2013).
19. Easton, J. D. *et al.* Definition and Evaluation of Transient Ischemic Attack: A Scientific Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Stroke Council; Council on Cardiovascular Surgery and Anesthesia; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Nursing; and the Interdisciplinary Council on Peripheral Vascular Disease: *The American*

Academy of Neurology affirms the value of this statement as an educational tool for neurologists. Stroke **40**, 2276–2293 (2009).

20. Mughal, M. M. *et al.* Symptomatic and asymptomatic carotid artery plaque. *Expert Rev. Cardiovasc. Ther.* **9**, 1315–1330 (2011).
21. Jashari, F. *et al.* Coronary and carotid atherosclerosis: Similarities and differences. *Atherosclerosis* **227**, 193–200 (2013).
22. Sethi, D., Gofur, E. M. & Munakomi, S. Anatomy, Head and Neck, Carotid Arteries. in *StatPearls* (StatPearls Publishing, 2020).
23. Hirschl, M. & Kundi, M. Carotid interventions and blood pressure. *Wien. Med. Wochenschr.* **164**, 503–507 (2014).
24. Kwon, T.-G. *et al.* Prevalence and Significance of Carotid Plaques in Patients With Coronary Atherosclerosis. *Korean Circ. J.* **39**, 317 (2009).
25. Touzé, E. Natural history of asymptomatic carotid artery stenosis. *Rev. Neurol. (Paris)* **164**, 793–800 (2008).
26. Flaherty, M. L. *et al.* Carotid Artery Stenosis as a Cause of Stroke. *Neuroepidemiology* **40**, 36–41 (2013).
27. Yang, J. *et al.* Wall shear stress in hypertensive patients is associated with carotid vascular deformation assessed by speckle tracking strain imaging. *Clin. Hypertens.* **20**, 10 (2014).
28. Gnasso, A. *et al.* In Vivo Association Between Low Wall Shear Stress and Plaque in Subjects With Asymmetrical Carotid Atherosclerosis. *Stroke* **28**, 993–998 (1997).
29. Medical gallery of Blausen Medical 2014. *WikiJournal Med.* **1**, (2014).
30. Yu, A. Y. X. & Coutts, S. B. Risk assessment to prevent recurrence after mild stroke or TIA. *Nat. Rev. Neurol.* **11**, 131–133 (2015).

31. Lovett, J. K. *et al.* Very Early Risk of Stroke After a First Transient Ischemic Attack. *Stroke* **34**, (2003).
32. Mohan, K. M. *et al.* Risk and Cumulative Risk of Stroke Recurrence: A Systematic Review and Meta-Analysis. *Stroke* **42**, 1489–1494 (2011).
33. Barnett, H. J. M. *et al.* Benefit of Carotid Endarterectomy in Patients with Symptomatic Moderate or Severe Stenosis. *N. Engl. J. Med.* **339**, 1415–1425 (1998).
34. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet Lond. Engl.* **351**, 1379–1387 (1998).
35. Rerkasem, K. & Rothwell, P. M. Carotid endarterectomy for symptomatic carotid stenosis. in *Cochrane Database of Systematic Reviews* (ed. The Cochrane Collaboration) CD001081.pub2 (John Wiley & Sons, Ltd, 2011). doi:10.1002/14651858.CD001081.pub2.
36. Brott, T. G. *et al.* Stenting versus Endarterectomy for Treatment of Carotid-Artery Stenosis. *N. Engl. J. Med.* **363**, 11–23 (2010).
37. Bonati, L. H. *et al.* Long-term outcomes after stenting versus endarterectomy for treatment of symptomatic carotid stenosis: the International Carotid Stenting Study (ICSS) randomised trial. *The Lancet* **385**, 529–538 (2015).
38. Spence, J. D., Song, H. & Cheng, G. Appropriate management of asymptomatic carotid stenosis. *BMJ* **1**, 64–71 (2016).
39. Prevention of disabling and fatal strokes by successful carotid endarterectomy in patients without recent neurological symptoms: randomised controlled trial. *The Lancet* **363**, 1491–1502 (2004).
40. Walker, M. D. Endarterectomy for Asymptomatic Carotid Artery Stenosis. *JAMA J. Am. Med. Assoc.* **273**, 1421 (1995).

41. Rudarakanchana, N., Dialynas, M. & Halliday, A. Asymptomatic Carotid Surgery Trial-2 (ACST-2): Rationale for a Randomised Clinical Trial Comparing Carotid Endarterectomy with Carotid Artery Stenting in Patients with Asymptomatic Carotid Artery Stenosis. *Eur. J. Vasc. Endovasc. Surg.* **38**, 239–242 (2009).
42. National Institute of Health and Care Excellence (NICE). Carotid artery stent placement for asymptomatic extracranial carotid stenosis: Interventional procedures guidance [IPG388]. https://www.nice.org.uk/guidance/ipg388/chapter/1-Guidance (2017).
43. Norris, J. W. & Zhu, C. Z. Stroke risk and critical carotid stenosis. *J. Neurol. Neurosurg. Psychiatry* **53**, 235–237 (1990).
44. Benavente, O., Moher, D. & Pham, B. Carotid endarterectomy for asymptomatic carotid stenosis: a meta-analysis. *BMJ* **317**, 1477–1480 (1998).
45. de Weerd, M. *et al.* Prevalence of Asymptomatic Carotid Artery Stenosis in the General Population: An Individual Participant Data Meta-Analysis. *Stroke* **41**, 1294–1297 (2010).
46. Magge, R. *et al.* Clinical Risk Factors and CT Imaging Features of Carotid Atherosclerotic Plaques as Predictors of New Incident Carotid Ischemic Stroke: A Retrospective Cohort Study. *Am. J. Neuroradiol.* **34**, 402–409 (2013).
47. Verhoeven, B. *et al.* Carotid atherosclerotic plaques in patients with transient ischemic attacks and stroke have unstable characteristics compared with plaques in asymptomatic and amaurosis fugax patients. *J. Vasc. Surg.* **42**, 1075–1081 (2005).
48. DeMarco, J. K. & Huston, J. Imaging of high-risk carotid artery plaques: current status and future directions. *Neurosurg. Focus* **36**, E1 (2014).
49. Gupta, A. *et al.* Carotid Plaque MRI and Stroke Risk: A Systematic Review and Meta-analysis. *Stroke* **44**, 3071–3077 (2013).
50. Saba, L. & Mallarini, G. MDCTA of Carotid Plaque Degree of Stenosis: Evaluation of Interobserver Agreement. *Am. J. Roentgenol.* **190**, W41–W46 (2008).

51. Das, M. *et al.* Carotid Plaque Analysis: Comparison of Dual-Source Computed Tomography (CT) Findings and Histopathological Correlation. *Eur. J. Vasc. Endovasc. Surg.* **38**, 14–19 (2009).
52. Kesavadas, C., Sylaja, P., Thomas, B., Kapilamoorthy, T. & Hingwala, D. Multimodality imaging of carotid atherosclerotic plaque: Going beyond stenosis. *Indian J. Radiol. Imaging* **23**, 26 (2013).
53. Forjoe, T. & Asad Rahi, M. Systematic review of preoperative carotid duplex ultrasound compared with computed tomography carotid angiography for carotid endarterectomy. *Ann. R. Coll. Surg. Engl.* **101**, 141–149 (2019).
54. Agatston, A. S. *et al.* Quantification of coronary artery calcium using ultrafast computed tomography. *J. Am. Coll. Cardiol.* **15**, 827–832 (1990).
55. Hecht, H. S. Coronary Artery Calcium Scanning. *JACC Cardiovasc. Imaging* **8**, 579–596 (2015).
56. Martin, S. S. *et al.* Dyslipidemia, Coronary Artery Calcium, and Incident Atherosclerotic Cardiovascular Disease: Implications for Statin Therapy From the Multi-Ethnic Study of Atherosclerosis. *Circulation* **129**, 77–86 (2014).
57. Using Nontraditional Risk Factors in Coronary Heart Disease Risk Assessment: U.S. Preventive Services Task Force Recommendation Statement. *Ann. Intern. Med.* **151**, 474 (2009).
58. Wang, X. *et al.* A zero coronary artery calcium score in patients with stable chest pain is associated with a good prognosis, despite risk of non-calcified plaques. *Open Heart* **6**, e000945 (2019).
59. Kwee, R. M. Systematic review on the association between calcification in carotid plaques and clinical ischemic symptoms. *J. Vasc. Surg.* **51**, 1015–1025 (2010).

60. Joshi, P. H. *et al.* Coronary artery Calcium predicts Cardiovascular events in participants with a low lifetime risk of Cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis* **246**, 367–373 (2016).
61. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006 (2014).
62. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577 (2016).
63. Rizzo, S. *et al.* CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. *Eur. Radiol.* **26**, 32–42 (2016).
64. Zhang, Y., Oikonomou, A., Wong, A., Haider, M. A. & Khalvati, F. Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer. *Sci. Rep.* **7**, 46349 (2017).
65. Kolossváry, M., Kellermayer, M., Merkely, B. & Maurovich-Horvat, P. Cardiac Computed Tomography Radiomics: A Comprehensive Review on Radiomic Techniques. *J. Thorac. Imaging* **33**, 26–34 (2018).
66. Prescott, J. W. Quantitative Imaging Biomarkers: The Application of Advanced Image Processing and Analysis to Clinical and Preclinical Decision Making. *J. Digit. Imaging* **26**, 97–108 (2013).
67. Ganeshan, B. *et al.* Non–Small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. *Radiology* **266**, 326–336 (2013).
68. Lubner, M. G., Smith, A. D., Sandrasegaran, K., Sahani, D. V. & Pickhardt, P. J. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *RadioGraphics* **37**, 1483–1503 (2017).
69. Liu, Y.-Z., Wang, Y.-X. & Jiang, C.-L. Inflammation: The Common Pathway of Stress-Related Diseases. *Front. Hum. Neurosci.* **11**, 316 (2017).

70. Sluimer, J. C. & Daemen, M. J. Novel concepts in atherogenesis: angiogenesis and hypoxia in atherosclerosis. *J. Pathol.* **218**, 7–29 (2009).
71. Haider, S. P., Burtness, B., Yarbrough, W. G. & Payabvash, S. Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas. *Cancers Head Neck* **5**, 6 (2020).
72. Ranjbar, S. & Ross Mitchell, J. An Introduction to Radiomics: An Evolving Cornerstone of Precision Medicine. in *Biomedical Texture Analysis* 223–245 (Elsevier, 2017). doi:10.1016/B978-0-12-812133-7.00008-9.
73. van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H. & Baessler, B. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging* **11**, 91 (2020).
74. Ganeshan, B. *et al.* CT-based texture analysis potentially provides prognostic information complementary to interim fdg-pet for patients with hodgkin's and aggressive non-hodgkin's lymphomas. *Eur. Radiol.* **27**, 1012–1020 (2017).
75. Chong, Y. *et al.* Quantitative CT Variables Enabling Response Prediction in Neoadjuvant Therapy with EGFR-TKIs: Are They Different from Those in Neoadjuvant Concurrent Chemoradiotherapy? *PLoS ONE* **9**, e88598 (2014).
76. Miles, K. A., Ganeshan, B. & Hayball, M. P. CT texture analysis using the filtration-histogram method: what do the measurements mean? *Cancer Imaging* **13**, 400–406 (2013).
77. Ganeshan, B. Quantitative Imaging Biomarkers from PET-CT as potential correlates for angiogenesis and hypoxia in colorectal cancer. (2012) doi:10.1594/ECR2012/B-0876.
78. Skogen, K., Ganeshan, B., Good, C., Critchley, G. & Miles, K. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. *J. Neurooncol.* **111**, 213–219 (2013).

79. Hu, Y. *et al.* Texture Feature Extraction and Analysis for Polyp Differentiation via Computed Tomography Colonography. *IEEE Trans. Med. Imaging* **35**, 1522–1531 (2016).

80. van Griethuysen, J. J. M. *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **77**, e104–e107 (2017).

81. Zwanenburg, A. *et al.* The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **295**, 328–338 (2020).

82. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973).

83. Haralick, R. M. Statistical and structural approaches to texture. *Proc. IEEE* **67**, 786–804 (1979).

84. Galloway, M. M. Texture analysis using gray level run lengths. *Comput. Graph. Image Process.* **4**, 172–179 (1975).

85. Thibault, G., Angulo, J. & Meyer, F. Advanced Statistical Matrices for Texture Characterization: Application to Cell Classification. *IEEE Trans. Biomed. Eng.* **61**, 630–637 (2014).

86. Sun, C. & Wee, W. G. Neighboring gray level dependence matrix for texture classification. *Comput. Vis. Graph. Image Process.* **23**, 341–352 (1983).

87. Amadasun, M. & King, R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man Cybern.* **19**, 1264–1274 (1989).

88. Buch, K. *et al.* Using Texture Analysis to Determine Human Papillomavirus Status of Oropharyngeal Squamous Cell Carcinomas on CT. *Am. J. Neuroradiol.* **36**, 1343–1348 (2015).

89. Ajana, S. *et al.* Benefits of dimension reduction in penalized regression methods for high-dimensional grouped data: a case study in low sample size. *Bioinformatics* **35**, 3628–3634 (2019).

90. Yang, M. *et al.* Imaging phenotype using radiomics to predict dry pleural dissemination in non-small cell lung cancer. *Ann. Transl. Med.* **7**, 259–259 (2019).

91. Kaissis, G. A. *et al.* Image-Based Molecular Phenotyping of Pancreatic Ductal Adenocarcinoma. *J. Clin. Med.* **9**, 724 (2020).

92. Ekert, K., Hinterleitner, C. & Horger, M. Prognosis assessment in metastatic gastrointestinal stromal tumors treated with tyrosine kinase inhibitors based on CT-texture analysis. *Eur. J. Radiol.* **116**, 98–105 (2019).

93. Vargas, H. A. *et al.* Radiogenomics of High-Grade Serous Ovarian Cancer: Multireader Multi-Institutional Study from the Cancer Genome Atlas Ovarian Cancer Imaging Research Group. *Radiology* **285**, 482–492 (2017).

94. Lo Gullo, R., Daimiel, I., Morris, E. A. & Pinker, K. Combining molecular and imaging metrics in cancer: radiogenomics. *Insights Imaging* **11**, 1 (2020).

95. Liu, Q. *et al.* Prediction of Aneurysm Stability Using a Machine Learning Model Based on PyRadiomics-Derived Morphological Features. *Stroke* **50**, 2314–2321 (2019).

96. Dolotova, D. *et al.* Application of Radiomics in Vesselness Analysis of CT Angiography Images of Stroke Patients. *Stud. Health Technol. Inform.* **270**, 33–37 (2020).

97. Oikonomou, E. K. *et al.* A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary CT angiography. *Eur. Heart J.* **40**, 3529–3543 (2019).

98. Christodoulou, C. I., Pattichis, C. S., Pantziaris, M. & Nicolaides, A. Texture-based classification of atherosclerotic carotid plaques. *IEEE Trans. Med. Imaging* **22**, 902–912 (2003).

99. Awad, J., Krasinski, A., Parraga, G. & Fenster, A. Texture analysis of carotid artery atherosclerosis from three-dimensional ultrasound images: Texture analysis of carotid artery atherosclerosis. *Med. Phys.* **37**, 1382–1391 (2010).

100. Larroza, A. *et al.* Texture analysis of cardiac cine magnetic resonance imaging to detect nonviable segments in patients with chronic myocardial infarction. *Med. Phys.* **45**, 1471–1480 (2018).

101. Larroza, A. *et al.* Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging. *Eur. J. Radiol.* **92**, 78–83 (2017).

102. Larroza, A., Lopez-Lereu, M. P., Monmeneu, J. V., Bodi, V. & Moratal, D. Texture analysis for infarcted myocardium detection on delayed enhancement MRI. in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* 1066–1069 (IEEE, 2017). doi:10.1109/ISBI.2017.7950700.

103. Neisius, U. *et al.* Radiomic Analysis of Myocardial Native T1 Imaging Discriminates Between Hypertensive Heart Disease and Hypertrophic Cardiomyopathy. *JACC Cardiovasc. Imaging* **12**, 1946–1954 (2019).

104. Otsuka, K. *et al.* Napkin-Ring Sign on Coronary CT Angiography for the Prediction of Acute Coronary Syndrome. *JACC Cardiovasc. Imaging* **6**, 448–457 (2013).

105. Kolossváry, M. *et al.* Radiomic Features Are Superior to Conventional Quantitative Computed Tomographic Metrics to Identify Coronary Plaques With Napkin-Ring Sign. *Circ. Cardiovasc. Imaging* **10**, (2017).

106. Kolossváry, M. *et al.* Identification of invasive and radionuclide imaging markers of coronary plaque vulnerability using radiomic analysis of coronary computed tomography angiography. *Eur. Heart J. - Cardiovasc. Imaging* **20**, 1250–1258 (2019).

107. Nam, K. *et al.* Value of Computed Tomography Radiomic Features for Differentiation of Periprosthetic Mass in Patients With Suspected Prosthetic Valve Obstruction. *Circ. Cardiovasc. Imaging* **12**, (2019).

108. Samuel, A. L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **3**, 210–229 (1959).

109. Mitchell, T. M., Keller, R. M. & Kedar-Cabelli, S. T. Explanation-Based Generalization: A Unifying View. *Mach. Learn.* **1**, 47–80 (1986).

110. Wu, T. H., Pang, G. K.-H. & Kwong, E. W.-Y. Predicting Systolic Blood Pressure Using Machine Learning. in *7th International Conference on Information and Automation for Sustainability* 1–6 (IEEE, 2014). doi:10.1109/ICIAFS.2014.7069529.

111. Wang, S. & Summers, R. M. Machine learning and radiology. *Med. Image Anal.* **16**, 933–951 (2012).

112. Dori, D. *Model-Based Systems Engineering with OPM and SysML*. (Springer, 2016).

113. Yassin, N. I. R., Omran, S., El Houby, E. M. F. & Allam, H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Comput. Methods Programs Biomed.* **156**, 25–45 (2018).

114. Le, E. P. V., Wang, Y., Huang, Y., Hickman, S. & Gilbert, F. J. Artificial intelligence in breast imaging. *Clin. Radiol.* **74**, 357–366 (2019).

115. Menze, B. H. *et al.* The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2015).

116. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).

117. Erickson, B. J. Deep Learning and Machine Learning in Imaging: Basic Principles. in *Artificial Intelligence in Medical Imaging* (eds. Ranschaert, E. R., Morozov, S. & Algra, P. R.) 39–46 (Springer International Publishing, 2019). doi:10.1007/978-3-319-94878-2_4.

118. Pelillo, M. Alhazen and the nearest neighbor rule. *Pattern Recognit. Lett.* **38**, 34–37 (2014).

119. Peng, C.-Y. J., Lee, K. L. & Ingersoll, G. M. An Introduction to Logistic Regression Analysis and Reporting. *J. Educ. Res.* **96**, 3–14 (2002).

120. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).

121. Svetnik, V. *et al.* Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **43**, 1947–1958 (2003).

122. Abiodun, O. I. *et al.* State-of-the-art in artificial neural network applications: A survey. *Heliyon* **4**, e00938 (2018).

123. Atienza, R. *Advanced deep learning with Keras apply deep learning techniques, autoencoders, GANs, variational autoencoders, deep reinforcement learning, policy gradients, and more*. (2018).

124. Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M. & Qureshi, N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLOS ONE* **12**, e0174944 (2017).

125. Goff, D. C. *et al.* 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* **129**, S49–S73 (2014).

126. Ambale-Venkatesh, B. *et al.* Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ. Res.* **121**, 1092–1101 (2017).

127. Cruz Rivera, S., Liu, X., Chan, A.-W., Denniston, A. K. & Calvert, M. J. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat. Med.* **26**, 1351–1363 (2020).

128. The SPIRIT-AI and CONSORT-AI Working Group *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat. Med.* **26**, 1364–1374 (2020).

129. Sengupta, P. P. *et al.* Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist. *JACC Cardiovasc. Imaging* **13**, 2017–2035 (2020).

130. Motwani, M. *et al.* Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur. Heart J.* ehw188 (2016) doi:10.1093/eurheartj/ehw188.

131. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLOS ONE* **14**, e0224365 (2019).

132. Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).

133. CS231n Convolutional Neural Networks for Visual Recognition. https://cs231n.github.io/neural-networks-3/.

134. Chlis, N. K. Comparison of Statistical Methods for Genomic Signature Extraction. (2013) doi:10.13140/2.1.2230.6563.

135. Kuhn, M. & Johnson, K. *Applied predictive modeling*. (Springer, 2013).

136. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).

137. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **61**, 85–117 (2015).

138. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

139. ter Haar Romeny, B. M. A Deeper Understanding of Deep Learning. in *Artificial Intelligence in Medical Imaging* (eds. Ranschaert, E. R., Morozov, S. & Algra, P. R.) 25–38 (Springer International Publishing, 2019). doi:10.1007/978-3-319-94878-2_3.

140. Kluger, J. *Simplexity: why simple things become complex (and how complex things can be made simple)*. (Hyperion, 2008).

141. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning*. (The MIT Press, 2016).

142. Chartrand, G. *et al.* Deep Learning: A Primer for Radiologists. *RadioGraphics* **37**, 2113–2131 (2017).

143. Mookiah, M. R. K. *et al.* Computer-aided diagnosis of diabetic retinopathy: A review. *Comput. Biol. Med.* **43**, 2136–2155 (2013).

144. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).

145. Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).

146. Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**, 2402 (2016).

147. Litjens, G. *et al.* State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC Cardiovasc. Imaging* **12**, 1549–1565 (2019).

148. McBee, M. P. *et al.* Deep Learning in Radiology. *Acad. Radiol.* **25**, 1472–1480 (2018).

149. Kumamaru, K. K., Hoppel, B. E., Mather, R. T. & Rybicki, F. J. CT Angiography: Current Technology and Clinical Use. *Radiol. Clin. North Am.* **48**, 213–235 (2010).

150. Hayashi, N. *et al.* Computed Tomography Demonstration of the Production and Distribution of Oxygen Gas Following Intratumoral Injection of a New Radiosensitizer (KORTUC) for Patients with Breast Cancer—Is Intratumoral Injection Not an Ideal Approach to Solve the Major Problem of Tumor Hypoxia in Radiotherapy? *Cancers* **8**, 43 (2016).

151. Evans, N. R. *et al.* Dual-Tracer Positron-Emission Tomography for Identification of Culprit Carotid Plaques and Pathophysiology In Vivo. *Circ. Cardiovasc. Imaging* **13**, (2020).

152. Tarkin, J. M. *et al.* Detection of Atherosclerotic Inflammation by 68 Ga-DOTATATE PET Compared to [18 F]FDG PET Imaging. *J. Am. Coll. Cardiol.* **69**, 1774–1791 (2017).

153. Joshi, F. R. *et al.* Vascular Imaging With 18 F-Fluorodeoxyglucose Positron Emission Tomography Is Influenced by Hypoxia. *J. Am. Coll. Cardiol.* **69**, 1873–1874 (2017).

154. Almutairi, A., Sun, Z., Poovathumkadavi, A. & Assar, T. Dual energy CT angiography of peripheral arterial disease: Feasibility of using lower contrast medium volume. *PLoS ONE* **10**, e0139275 (2015).

155. Guziński, M., Waszczuk, Ł. & Sąsiadek, M. J. Head CT: Image quality improvement of posterior fossa and radiation dose reduction with ASiR - comparative studies of CT head examinations. *Eur. Radiol.* **26**, 3691–3696 (2016).

156. Engel, L. C. *et al.* Ultra-Low Dose Cardiac CT Angiography at 80 kV using Second Generation Dual-Source CT: Assessment of Radiation Dose and Image Quality. *J. Med. Diagn. Methods* **01**, 1–7 (2012).

157. Motoyama, S. *et al.* Plaque characterization by coronary computed tomography angiography and the likelihood of acute coronary events in mid-term follow-up. *J. Am. Coll. Cardiol.* **66**, 337–346 (2015).

158. North American Symptomatic Carotid Endarterectomy Trial. Methods, patient characteristics, and progress. *Stroke* **22**, (1991).

159. Wang, C. *et al.* Building CT Radiomics-Based Models for Preoperatively Predicting Malignant Potential and Mitotic Count of Gastrointestinal Stromal Tumors. *Transl. Oncol.* **12**, 1229–1236 (2019).

160. Krejza, J. *et al.* Carotid Artery Diameter in Men and Women and the Relation to Body and Neck Size. *Stroke* **37**, 1103–1105 (2006).

161. Zaccagna, Fulvio. CT Texture of Carotid Arteries identifies Vulnerable Plaque in Stroke and Transient Ischaemic Attack: A Preliminary Outcome Study. (2017).

162. Ibrahim, A. *et al.* Radiomics Analysis for Clinical Decision Support in Nuclear Medicine. *Semin. Nucl. Med.* **49**, 438–449 (2019).

163. De Cecco, C. N. *et al.* Texture Analysis as Imaging Biomarker of Tumoral Response to Neoadjuvant Chemoradiotherapy in Rectal Cancer Patients Studied with 3-T Magnetic Resonance: *Invest. Radiol.* **50**, 239–245 (2015).

164. Ng, F., Kozarski, R., Ganeshan, B. & Goh, V. Assessment of tumor heterogeneity by CT texture analysis: Can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur. J. Radiol.* **82**, 342–348 (2013).

165. U-King-Im, J. M. K. S. *et al.* Measuring Carotid Stenosis on Contrast-Enhanced Magnetic Resonance Angiography: Diagnostic Performance and Reproducibility of 3 Different Methods. *Stroke* **35**, 2083–2088 (2004).

166. National Institute of Health and Care Excellence (NICE). Stroke and transient ischaemic attack in over 16s: diagnosis and initial management (NICE Guideline NG128). https://www.nice.org.uk/guidance/ng128/chapter/Recommendations (2019).

167. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155–163 (2016).

168. Cattell, R., Chen, S. & Huang, C. Robustness of radiomic features in magnetic resonance imaging: review and a phantom study. *Vis. Comput. Ind. Biomed. Art* **2**, 19 (2019).

169. Larue, R. T. H. M., Defraene, G., De Ruysscher, D., Lambin, P. & van Elmpt, W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br. J. Radiol.* **90**, 20160665 (2017).

170. Ng, F., Ganeshan, B., Kozarski, R., Miles, K. A. & Goh, V. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: Contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology* **266**, 177–184 (2013).

171. Dercle, L. *et al.* Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: ROI-area, acquisition protocol and tissue site exert substantial influence. *Sci. Rep.* **7**, 7952 (2017).

172. Deng, Y. *et al.* CT texture analysis in the differentiation of major renal cell carcinoma subtypes and correlation with Fuhrman grade. *Eur. Radiol.* **29**, 6922–6929 (2019).

173. Van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).

174. Fiset, S. *et al.* Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother. Oncol.* **135**, 107–114 (2019).

175. on, M. M. DICOM Rescale Intercept / Rescale Slope and ITK. *Kitware Blog* https://blog.kitware.com/dicom-rescale-intercept-rescale-slope-and-itk/ (2014).

176. Lorensen, W. E. & Cline, H. E. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Comput. Graph.* **21**, 163–169 (1987).

177. Le, E. P. V. *et al.* 105 Machine learning and carotid artery CT radiomics identify significant differences between culprit and non-culprit lesions in patients with stroke and transient ischaemic attack. *Heart* **106**, A82 (2020).

178. Image processing — IBSI 0.0.1dev documentation. https://ibsi.readthedocs.io/en/latest/02_Image_processing.html.

179. de Weert, T. T. *et al.* Assessment of atherosclerotic carotid plaque volume with multidetector computed tomography angiography. *Int. J. Cardiovasc. Imaging* **24**, 751–759 (2008).

180. Kumar, R. M. A Survey on Image Feature Descriptors. **5**, 6 (2014).

181. Edalat-Javid, M. *et al.* Cardiac SPECT radiomic features repeatability and reproducibility: A multi-scanner phantom study. *J. Nucl. Cardiol.* (2020) doi:10.1007/s12350-020-02109-0.

182. Zwanenburg, A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur. J. Nucl. Med. Mol. Imaging* **46**, 2638–2655 (2019).

183. Mackin, D. *et al.* Measuring Computed Tomography Scanner Variability of Radiomics Features: *Invest. Radiol.* **50**, 757–765 (2015).

184. Hatt, M., Tixier, F., Cheze Le Rest, C., Pradier, O. & Visvikis, D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur. J. Nucl. Med. Mol. Imaging* **40**, 1662–1671 (2013).

185. Jang, J. *et al.* Reproducibility of Myocardial Radiomic Features in Cardiac MRI Imaging. *Radiol. Cardiothorac. Imaging* (2020).

186. Zwanenburg, A. *et al.* Assessing robustness of radiomic features by image perturbation. *Sci. Rep.* **9**, 614 (2019).

187. Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index1. *Acad. Radiol.* **11**, 178–189 (2004).

188. Frequently Asked Questions — pyradiomics v3.0.post5+gf06ac1d documentation. https://pyradiomics.readthedocs.io/en/latest/faq.html?highlight=bin%20width#what-about-gray-value-discretization-fixed-bin-width-fixed-bin-count.

189. Forghani, R. *et al.* Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Comput. Struct. Biotechnol. J.* **17**, 995–1008 (2019).

190. Feng, B. *et al.* Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *Eur. J. Radiol.* **128**, 109022 (2020).

191. Owens, C. A. *et al.* Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLOS ONE* **13**, e0205003 (2018).

192. Shafiq-ul-Hassan, M. *et al.* Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci. Rep.* **8**, 10545 (2018).

193. Moore, D. H. Classification and regression trees, by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. Brooks/Cole Publishing, Monterey, 1984,358 pages, $27.95. *Cytometry* **8**, 534–535 (1987).

194. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 (IEEE Comput. Soc. Press, 1995).

195. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).

196. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).

197. Aitkin, M. & Foxall, R. Statistical modelling of artificial neural networks using the multi-layer perceptron. *Stat. Comput.* **13**, 227–239 (2003).

198. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016). doi:10.1145/2939672.2939785.

199. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).

200. Leijenaar, R. T. H. *et al.* The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci. Rep.* **5**, 11075 (2015).

201. Da-ano, R. *et al.* Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci. Rep.* **10**, 10248 (2020).

202. Rudd, J. H. F. *et al.* Imaging Atherosclerotic Plaque Inflammation With [^{18} F]-Fluorodeoxyglucose Positron Emission Tomography. *Circulation* **105**, 2708–2711 (2002).

203. Chowdhury, M. M. *et al.* 18 F-FDG Uptake on PET/CT in Symptomatic versus Asymptomatic Carotid Disease: a Meta-Analysis. *Eur. J. Vasc. Endovasc. Surg.* **56**, 172–179 (2018).

204. Sriranjan, R. S. *et al.* Atherosclerosis imaging using PET: Insights and applications. *Br. J. Pharmacol.* bph.14868 (2019) doi:10.1111/bph.14868.

205. Tawakol, A. *et al.* Intensification of Statin Therapy Results in a Rapid Reduction in Atherosclerotic Inflammation. *J. Am. Coll. Cardiol.* **62**, 909–917 (2013).

206. Evans, N. Multimodal Imaging of Inflammation at the Neurovascular Interface in Cerebrovascular Disease. (University of Cambridge, 2017).

207. Lall, S. & Boyd, S. Validation EE104 Stanford University.

208. (17) How do you measure specificity and sensitivity in a multiple class classification problem? *ResearchGate* https://www.researchgate.net/post/How-do-you-measure-specificity-and-sensitivity-in-a-multiple-class-classification-problem.

209. Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).

210. Molnar, C. *Interpretable machine learning*. (2019).

211. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems 30* (eds. Guyon, I. et al.) 4765–4774 (Curran Associates, Inc., 2017).

212. Sanduleanu, S. *et al.* Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiother. Oncol.* **127**, 349–360 (2018).

213. Tarkin, J. M. Atherosclerotic inflammation imaging using somatostatin receptor-2 positron emission tomography. (2017) doi:10.17863/CAM.15485.

214. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

215. Jensen, E. C. Quantitative Analysis of Histological Staining and Fluorescence Using ImageJ: Histological Staining/Fluorescence Using ImageJ. *Anat. Rec.* **296**, 378–381 (2013).

216. Shen, Matthew. How can I calculate % area stained using ImageJ and IHC toolbox? https://www.researchgate.net/post/How_I_can_calculate_area_stained_using_ImageJ_and_IHC_toolbox/ (2019).

217. Ruifrok, A. C. & Johnston, D. A. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**, 291–299 (2001).

218. Tsanas, A., Little, M. A. & McSharry, P. E. A Methodology for the Analysis of Medical Data. in *Handbook of Systems and Complexity in Health* (eds. Sturmberg, J. P. & Martin, C. M.) 113–125 (Springer New York, 2013). doi:10.1007/978-1-4614-4998-0_7.

219. Omar, A., Chatterjee, T. K., Tang, Y., Hui, D. Y. & Weintraub, N. L. Proinflammatory Phenotype of Perivascular Adipocytes. *Arterioscler. Thromb. Vasc. Biol.* **34**, 1631–1636 (2014).

220. Lin, A. *et al.* Myocardial Infarction Associates With a Distinct Pericoronary Adipose Tissue Radiomic Phenotype. *JACC Cardiovasc. Imaging* S1936878X2030615X (2020) doi:10.1016/j.jcmg.2020.06.033.

221. Ho, Y. C. & Pepyne, D. L. Simple Explanation of the No-Free-Lunch Theorem and Its Implications. *J. Optim. Theory Appl.* **115**, 549–570 (2002).

222. Shiri, I. *et al.* Next-Generation Radiogenomics Sequencing for Prediction of EGFR and KRAS Mutation Status in NSCLC Patients Using Multimodal Imaging and Machine Learning Algorithms. *Mol. Imaging Biol.* **22**, 1132–1148 (2020).

223. Kocak, B. *et al.* Radiomics with artificial intelligence: a practical guide for beginners. *Diagn. Interv. Radiol.* **25**, 485–495 (2019).

224. Nerlekar, N. *et al.* Computed Tomographic Coronary Angiography–Derived Plaque Characteristics Predict Major Adverse Cardiovascular Events: A Systematic Review and Meta-Analysis. *Circ. Cardiovasc. Imaging* **11**, (2018).

225. Ganeshan, B., Abaleke, S., Young, R. C. D., Chatwin, C. R. & Miles, K. A. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging* **10**, 137–143 (2010).

226. Xu, Z., Li, X.-F., Zou, H., Sun, X. & Shen, B. 18F-Fluoromisonidazole in tumor hypoxia imaging. *Oncotarget* **8**, 94969–94979 (2017).

227. Huang, J. *et al.* DeepPrimitive: Image decomposition by layered primitive detection. *Comput. Vis. Media* **4**, 385–397 (2018).

228. Gill, A. B. *et al.* Correlating Radiomic Features of Heterogeneity on CT with Circulating Tumor DNA in Metastatic Melanoma. *Cancers* **12**, 3493 (2020).

229. Arcuri, A. & Fraser, G. Parameter tuning or default values? An empirical investigation in search-based software engineering. *Empir. Softw. Eng.* **18**, 594–623 (2013).

230. Probst, P., Boulesteix, A.-L. & Bischl, B. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. 32.

231. Shultz, T. R. *et al.* Curse of Dimensionality. in *Encyclopedia of Machine Learning* (eds. Sammut, C. & Webb, G. I.) 257–258 (Springer US, 2011). doi:10.1007/978-0-387-30164-8_192.

232. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **2**, 303–314 (1989).

233. Gao, X. & Qian, Y. Application of Deep Learning Neural Network for Classification of TB Lung CT Images Based on Patches. *CEUR Workshop Proc.* 9 (2017).

234. Han, J. & Kamber, M. *Data mining: concepts and techniques*. (Elsevier, 2012).

235. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs* (2015).

236. Baum, E. & Wilczek, F. Supervised Learning of Probability Distributions by Neural Networks. in *Neural Information Processing Systems* (ed. Anderson, D.) 52–61 (American Institute of Physics, 1988).

237. Chollet, F. *Deep learning with Python*. (Manning Publications Co, 2018).

238. Lin, M., Chen, Q. & Yan, S. Network In Network. *ArXiv13124400 Cs* (2014).

239. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning Deep Features for Discriminative Localization. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2921–2929 (IEEE, 2016). doi:10.1109/CVPR.2016.319.

240. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv150203167 Cs* (2015).

241. Santurkar, S., Tsipras, D., Ilyas, A. & Madry, A. How Does Batch Normalization Help Optimization? in *Advances in Neural Information Processing Systems 31* (eds. Bengio, S. et al.) 2483–2493 (Curran Associates, Inc., 2018).

242. Brownlee, J. Dropout Regularization in Deep Learning Models With Keras. *Machine Learning Mastery* https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/ (2016).

243. Liu, Y., Chen, P.-H. C., Krause, J. & Peng, L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *JAMA* **322**, 1806 (2019).

244. Alcantara, G. Empirical analysis of non-linear activation functions for Deep Neural Networks in classification tasks. *ArXiv171011272 Cs Stat* (2017).

245. Neishi, M. *et al.* A Bag of Useful Tricks for Practical Neural Machine Translation: Embedding Layer Initialization and Large Batch Size. in *Proceedings of the 4th Workshop on Asian Translation (WAT2017)* 99–109 (Asian Federation of Natural Language Processing, 2017).

246. You, K., Long, M., Wang, J. & Jordan, M. I. How Does Learning Rate Decay Help Modern Neural Networks? (2019).

247. Ruder, S. An overview of gradient descent optimization algorithms. *ArXiv160904747 Cs* (2017).

248. Hinton, G., Srivastava, N. & Swersky, K. Neural Networks for Machine Learning: Lecture 6a Overview of mini-batch gradient descent. (2014).

249. (Tutorial) Convolutional Neural Networks with TensorFlow. *DataCamp Community* https://www.datacamp.com/community/tutorials/cnn-tensorflow-python (2020).

250. Initializers - Keras 2.0.6. Documentation. https://faroit.com/keras-docs/2.0.6/initializers/#glorot_uniform.

251. General Data Protection Regulation (GDPR) – Official Legal Text. *General Data Protection Regulation (GDPR)* https://gdpr-info.eu/.

252. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **128**, 336–359 (2020).

253. Alpha Blending using OpenCV (C++ / Python) | Learn OpenCV. https://www.learnopencv.com/alpha-blending-using-opencv-cpp-python/.

254. Szegedy, C. *et al.* Intriguing properties of neural networks. *ArXiv13126199 Cs* (2014).

255. De Fauw, J. *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).

256. Caruana, R. Multitask Learning. *Mach. Learn.* **28**, 41–75 (1997).

257. Yang, Q., Liu, Y., Chen, T. & Tong, Y. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* **10**, 1–19 (2019).

258. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).

259. Sutskever, I., Martens, J., Dahl, G. & Hinton, G. On the importance of initialization and momentum in deep learning. 14.

260. Lambin, P. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).

261. Avanzo, M., Stancanello, J. & El Naqa, I. Beyond imaging: The promise of radiomics. *Phys. Med.* **38**, 122–139 (2017).

262. Gu, Y. *et al.* Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recognit.* **46**, 692–702 (2013).

263. Kotanidis, C. P. *et al.* Pericoronary fat radiomic profile (FRP) predicts long-term cardiac risk in individuals with calcium score below 100 on coronary computed tomography angiography. *Eur. Heart J.* **41**, ehaa946.0181 (2020).

264. Glocker, B., Robinson, R., Castro, D. C., Dou, Q. & Konukoglu, E. Machine Learning with Multi-Site Imaging Data: An Empirical Study on the Impact of Scanner Effects. *ArXiv191004597 Cs Eess Q-Bio* (2019).

265. Bera, K., Velcheti, V. & Madabhushi, A. Novel Quantitative Imaging for Predicting Response to Therapy: Techniques and Clinical Applications. *Am. Soc. Clin. Oncol. Educ. Book* 1008–1018 (2018) doi:10.1200/EDBK_199747.

266. Pérez-Medina, C., Fayad, Z. A. & Mulder, W. J. M. Atherosclerosis Immunoimaging by Positron Emission Tomography. *Arterioscler. Thromb. Vasc. Biol.* **40**, 865–873 (2020).

267. Ganeshan, B. & Miles, K. A. Quantifying tumour heterogeneity with CT. *Cancer Imaging* **13**, 140–149 (2013).

268. Lv, W. *et al.* Radiomics Analysis of PET and CT Components of PET/CT Imaging Integrated with Clinical Parameters: Application to Prognosis for Nasopharyngeal Carcinoma. *Mol. Imaging Biol.* **21**, 954–964 (2019).

269. Sala, E. *et al.* Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin. Radiol.* **72**, 3–10 (2017).

270. Berger, M. F. & Mardis, E. R. The emerging clinical relevance of genomics in cancer medicine. *Nat. Rev. Clin. Oncol.* **15**, 353–365 (2018).

271. Smith, S. M. *et al.* Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *eLife* **9**, e52677 (2020).

272. Berman, S. E. *et al.* The relationship between carotid artery plaque stability and white matter ischemic injury. *NeuroImage Clin.* **9**, 216–222 (2015).

Appendix: Prizes, Papers and Presentations during PhD

Prizes

2020	Highly Commended for the Master Surgeon International Essay Prize
2020	Highly Commended Virtual Abstract at the International Collaborative Grand Rounds (ICGRx) Summit
2020	Awarded Project Excellence with Code First Girls
2018-2019	Bertelsmann Data Science Challenge Scholarship
2018	1st place ARM hardware prize and 2nd place Microsoft Cognitive Vision Prize at the Cambridge hackathon

Publications

1. **Le EPV**, Rundo L, Tarkin JM, Evans NR, Chowdhury MM, Coughlin PA et al. Assessing Robustness of Carotid Artery CT Angiography Radiomics in the Identification of Culprit versus Non-culprit lesions in Cerebrovascular Events. *Manuscript under peer review, Nature Scientific Reports*
2. Utukuri M, D'Souza F, Deighton A, **Le EPV**, Osei-Boadu B, Gadi N et al. Digital Health: a neglected part of health curricula? *Manuscript under peer review, Digital Health*
3. Wall C, Huang Y, Uy C, **Le E**, Tombetti E, Gopalan D, et al. Pericoronary adipose tissue density is associated with clinical disease activity in Takayasu arteritis and coronary arterial inflammation measured by 68Ga-DOTATATE PET in atherosclerosis. Eur Heart J. 2020 Nov 1;41(Supplement_2):ehaa946.0182.
4. Gopalan D, Nordgren-Rogberg A, **Le EPV**, Pavey H, Tarkin J, Nyrén S, et al. Abnormal Pulmonary Venous Filling: An Adjunct Feature in the Computed Tomography Pulmonary Angiogram Assessment of Chronic Thromboembolic Pulmonary Hypertension. J Am Heart Assoc. 2020 Nov 3;9(21).
5. Evans NR, Tarkin JM, **Le EP**, Sriranjan RS, Corovic A, Warburton EA, et al. Integrated cardiovascular assessment of atherosclerosis using PET/MRI. Br J Radiol. 2020 Sep 1;93(1113):20190921.

6. Chowdhury MM, Tarkin JM, Albaghdadi MS, Evans NR, **Le EPV**, Berrett TB, et al. Vascular Positron Emission Tomography and Restenosis in Symptomatic Peripheral Arterial Disease. JACC Cardiovasc Imaging. 2020 Apr;13(4):1008–17.
7. Evans NR, Tarkin JM, Chowdhury MM, **Le EPV**, Coughlin PA, Rudd JHF, et al. Dual-Tracer Positron-Emission Tomography for Identification of Culprit Carotid Plaques and Pathophysiology In Vivo. Circ Cardiovasc Imaging. 2020 Mar;13(3).
8. Sriranjan RS, Tarkin JM, Evans NR, **Le EPV**, Chowdhury MM, Rudd JHF. Atherosclerosis imaging using PET: Insights and applications. Br J Pharmacol. 2019 Nov 14;bph.14868.
9. **Le EPV**, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol. 2019 May;74(5):357–66.
10. Wang X, **Le EPV**, Rajani NK, Hudson-Peacock N, Pavey H, Tarkin JM, et al. A zero coronary artery calcium score in patients with stable chest pain is associated with a good prognosis, despite risk of non-calcified plaques. Open Heart. 2019 Apr;6(1):e000945.
11. Tarkin JM, **Le EPV**, Calcagno C, Dweck MR, Evans NR, Chowdhury MM, et al. P30 68Ga-DOTATATE PET IDENTIFIES MYOCARDIAL INFLAMMATION AND BONE MARROW MONOCYTE MOBILISATION AFTER MYOCARDIAL INFARCTION. Cardiovasc Res. 2018 Sep 1;114(suppl_2):S9–10.
12. Chowdhury MM, Tarkin JM, Evans NR, **Le E**, Warburton EA, Hayes PD, et al. 18 F-FDG Uptake on PET/CT in Symptomatic versus Asymptomatic Carotid Disease: a Meta-Analysis. Eur J Vasc Endovasc Surg. 2018 Aug;56(2):172–9.

Conference presentations

1. Radiomics applied to carotid CT angiograms can identify significant differences between culprit and non-culprit lesions in patients with stroke and transient ischaemic attack – European Society of Cardiology (ESC) Congress 2020. Online due to COVID-19 pandemic. Subsequently published as an abstract in *European Heart Journal.*
    a. **Le E,** Evans NR, Tarkin JM, Chowdhury MM, Zaccagna F, Pavey H, et al. Radiomics applied to carotid CT angiograms can identify significant differences between culprit and non-culprit lesions in patients with stroke and transient ischaemic attack. Eur Heart J. 2020 Nov 1;41(Supplement_2):ehaa946.2417.
2. Contrast CT classification of asymptomatic and symptomatic carotids in stroke and transient ischaemic attack with deep learning and interpretability – European Society of Cardiology (ESC) Congress 2020. Online due to COVID-19 pandemic. Subsequently published as an abstract in *European Heart Journal.*
    a. **Le EPV,** Evans NR, Tarkin JM, Chowdhury MM, Zaccagna F, Wall C, et al. Contrast CT classification of asymptomatic and symptomatic carotids in stroke and transient ischaemic attack with deep learning and interpretability. Eur Heart J. 2020 Nov 1;41(Supplement_2):ehaa946.2418.
3. Machine learning and carotid artery CT radiomics identify significant differences between culprit and non-culprit lesions in patients with stroke and transient ischaemic attack. Published in HEART in lieu of the cancelled British Cardiovascular Society Annual Conference in Manchester, June 2020 due to COVID-19 pandemic.